# Journal of
# Experimental Psychology

# Volume 64, 1962

# CONTENTS OF VOLUME 64

## THE PARTIAL REINFORCEMENT EFFECT SUSTAINED THROUGH BLOCKS OF CONTINUOUS REINFORCEMENT [1]

JOHN THEIOS [2]

*Stanford University*

The discrimination interpretation of partial reinforcement (Bitterman, Fedderson, & Tyler, 1953; Mowrer & Jones, 1945) holds that resistance to extinction is a function of the degree of similarity between the training reinforcement schedule and the continuous nonreinforcement of extinction. According to this point of view, it is relatively easy for continuously reinforced $S$s to discriminate when extinction starts because of the sharp contrast in the reinforcing events at the onset of extinction. Partially reinforced $S$s have difficulty in discriminating the transition from training to extinction because they have experienced runs of nonreinforced trials, and the first few extinction trials provide a stimulus situation which is similar to that experienced during the nonreinforced runs in training.

In stimulus generalization terms, continuously reinforced $S$s experience a large stimulus change during extinction which results in a large decrement in response strength. Partially reinforced $S$s experience only a small stimulus change, and their response strength is decreased only slightly. The difference in resistance to extinction is attributable to the differential stimulus generalization decrements.

A possible test of the discrimination hypothesis would consist of varying the number of continuously reinforced trials interpolated between an initial partial reinforcement series and an extinction series. The continuously reinforced trials would serve to isolate the partial reinforcement training from extinction by permitting the partially reinforced $S$s to have a sharp contrast in the reinforcing events at the onset of extinction similar to that experienced by con-

1

tinuously reinforced $Ss$. This stimulus change experienced by $Ss$ that have received some continuous reinforcement following partial reinforcement should be larger than the change experienced by $Ss$ that received only partial reinforcement, and according to the discrimination hypothesis, the former $Ss$ should extinguish faster than the latter $Ss$. In general, the longer the block of continuous reinforcement given after partial reinforcement, the less similarity there should be between training and extinction. Therefore, resistance to extinction is expected to decrease with the number of continuously reinforced trials interpolated between partial reinforcement and extinction. As the number of interpolated trials becomes large, resistance to extinction should approach that of continuously reinforced $Ss$. The present experiment was designed to test these predictions from the discrimination theory of partial reinforcement.

## METHOD

*Apparatus.*—The apparatus was a black wooden runway 5 in. wide, 5 in. high, and 99 in. long. The first 12 in. comprised the start box and the last 15 in. the goal box. The start and goal boxes were separated from the alley by metal guillotine doors. A $1.8 \times 2.5$ in. food cup was concealed behind a 1.6-in. metal shield at the end of the goal box. The presence of the shield required $S$ to put his head into the food cup to see its contents.

Response time was measured in .01 sec. from the opening of the start door until $S$ interrupted a photobeam 5 in. inside the goal box. On rewarded trials a food pellet was automatically delivered by a solenoid feeder when $S$ interrupted a second photobeam positioned across the food cup. The reward consisted of one "Frostyos" pellet, a sugar-coated breakfast cereal produced by General Mills, Incorporated. On nonrewarded trials the solenoid operated, but no food was delivered. The $Ss$ required about 15 sec. to consume the pellet, and were removed to the

home cage 15 sec. after the interruption of the first photobeam.

*Subjects.*—The $Ss$ were 60 naive Slonaker albino rats, 28 males and 32 females, selected from the colony maintained by the Stanford University Psychology Department. They were approximately 90 days old at the beginning of the experiment and had been tamed for 3 days prior to the experiment. The $Ss$ were housed in individual cages with ad lib. water and were maintained on 22 hr. food deprivation throughout the experiment. The $Ss$ were permitted to eat lab chow for 2 hr. after each daily experimental session.

*Procedure.*—After taming, $S$ was placed into the goal box for 5 trials. Reward pellets were scattered throughout the goal box on the first 2 trials, but only the food cup contained pellets on the last 3 trials. When $S$ ate one pellet he was removed to his home cage. Preliminary runway training followed on the next 6 days at 5 trials a day. The 30 preliminary runway trials were continuously rewarded for all $Ss$. The $Ss$ were run in rotation, and the intertrial interval decreased from approximately 60 to 30 min. during this preliminary training.

After preliminary training, $Ss$ were divided into five groups, equated on performance during the preliminary 30 trials. During the remaining part of the experiment $Ss$ received 5 runway trials each day. Three groups of 14 $Ss$ each received 40% partial reinforcedment (P) for 70 trials. The schedule was random except for the restrictions that runs of longer than 6 rewarded or nonrewarded trials were excluded and that the series ended with a rewarded trial. One of the P groups was reduced to 13 $Ss$ when a rat died during the experiment. The remaining two groups, each consisting of 9 $Ss$, received continuous reinforcement (C) for 70 trials. All $Ss$ were left in the goal box for 15 sec. on each trial during this phase, irrespective of whether the trial was rewarded or not. The intertrial intervals averaged about 10 min.

Groups P-0 and C-0 were extinguished following the last of the 70 training trials. Group P-25, however, received 25 additional 100% rewarded trials, and then was extinguished. Groups P-70 and C-70 were run for 14 additional days on 100% reward, receiving 70 additional continuously rewarded trials before they were extinguished.

During extinction $S$ received 5 trials each day until he accumulated 5 trials, not necessarily consecutive, during which he did not enter the goal box within 60 sec. In order to

obtain extinction curves based on all $S$s, each $S$ was run for at least 40 trials. If $S$ did not enter the goal box within 60 sec., he was removed to his home cage and given a score of 60 sec. If $S$ entered the goal box, he was detained in the box for 15 sec. The empty feeder operated if $S$ looked into the food cup. The intertrial interval averaged about 10 min.

## RESULTS

*Training.*—Curves of response speed during training are presented in Fig. 1. An $S$'s median response speed for each block of 5 trials was computed. Each point on the curves represents a group mean of these median scores for a given block of trials. The $S$s were divided into five groups just before Block 7. An analysis of variance on the median speed scores of Block 7 yielded an $F < 1.0$. Hence, the division of $S$s did result in comparable groups.

The curves show that between Blocks 7 and 20, the trials on which the P $S$s received 40% reinforcement, the continuously reinforced $S$s eventually ran faster than the partially reinforced $S$s. The means and $SD$s of $S$s' median speeds during Block 20 are presented in Table 1. An analysis of variance performed on these scores indicated that there was no significant difference between the two continuously reinforced groups ($F = .28$, $df = 1/54$), nor among the three

### TABLE 1
MEANS AND $SD$s OF SPEED SCORES ON BLOCK 20

| Group | Mean | SD | N |
|-------|------|-----|----|
| P-0   | .64  | .07 | 13 |
| P-25  | .66  | .07 | 14 |
| P-70  | .62  | .08 | 14 |
| C-0   | .70  | .09 | 9  |
| C-70  | .72  | .11 | 9  |

partial groups ($F = 1.14$, $df = 2/54$). However, the partial groups ran significantly slower than the 100% groups ($F = 7.16$, $df = 1/54$, $P < .01$).

After Block 20, when Groups P-25 and P-70 were changed from 40% reinforcement to 100%, their response speeds increased essentially to that of the C-70 $S$s. Since Group P-25 was extinguished following Block 25, a comparison of the P groups and Group C-70 was made by an analysis of variance on the Block 25 median speed scores. There was no significant difference between the two P groups ($F = 1.13$, $df = 1/34$) nor between the combined P groups and Group C-70 ($F = 1.00$, $df = 1/34$). Thus, while on 40% reinforcement, the P $S$s ran significantly slower than the 100% $S$s. When the P $S$s were subsequently given 100% reinforcement, however, their response speeds significantly increased so that the initial difference between the 40% and 100% $S$s disappeared in less than 25 trials.

*Extinction.*—Curves of response time during extinction are given in Fig. 2. Each point on the curves represents a group mean of individual median time scores for a given block of 5 extinction trials. The striking feature of Fig. 2 is the difference between the extinction curves of the P and C groups. The three P groups, all of which initially received 40% reinforcement, have very similar extinc-
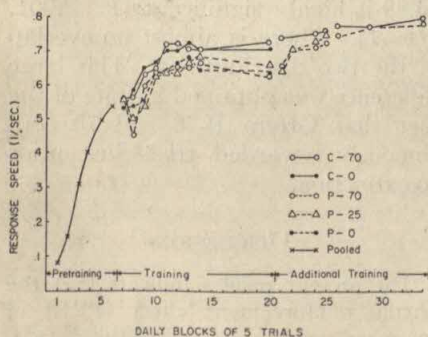


FIG. 1. Response speeds during training.

tion curves in spite of the fact that one group had 0, the second group 25, and the third 70 trials of 100% reinforcement between partial reinforcement and extinction. The curves of the two C groups are also very similar to each other. The difference between the P and C groups is evident as early as the second block of 5 extinction trials and indicates much faster extinction for the C groups. Thus, additional trials of 100% reinforcement following 40% reinforcement had little effect on rate of extinction.

The large difference in extinction rate between the P and C groups evidenced by the curves is also reflected in the mean number of trials to reach the extinction criterion (Table 2). All three P groups took at least 50 trials to reach the criterion, but the C groups took only about 32 trials. It can be seen that the $SD$s for the C groups are much smaller than those for the P groups, making inadvisable the use of an overall analysis of variance on the responses to the criterion scores. For Group C-0 and Group C-70, a $t$ of .78 was obtained, indicating that, for 100% reinforced $S$s, a difference between 100 and 170 training trials had no

TABLE 2

MEANS AND $SD$s OF RESPONSES TO THE EXTINCTION CRITERION

| Group | Mean | SD | N |
|-------|------|-----|----|
| P-0   | 61   | 14  | 13 |
| P-25  | 66   | 11  | 14 |
| P-70  | 50   | 10  | 14 |
| C-0   | 33   | 5   | 9  |
| C-70  | 31   | 6   | 9  |

effect on resistance to extinction. Homogeneity of variance among the responses to the criterion for the three P groups permitted a $1 \times 3$ analysis of variance. The obtained $F$ of 6.26 ($df = 2/38$) is significant at the .01 level, and indicates that individual $t$ tests are appropriate. Using the overall within-groups variance as the variance estimate for the error term of the $t$ tests, Groups P-0 and P-25 are not significantly different from each other ($t = .88$). The $t$ between Groups P-0 and P-70 of 2.46, however, is significant ($df = 25$, $P < .05$). Similarly, the $t$ between Groups P-25 and P-70 of 3.40 is significant ($df = 26$, $P < .01$).

In order to demonstrate that all the P groups are significantly more resistant to extinction than the C groups, a median test was run between Group P-70, the least resistant of the P groups, and the combination of the two C groups. The obtained $\chi^2$ of 24.89 is highly significant ($P < .001$, $df = 1$). There is almost no overlap in the two distributions. This large difference was obtained in spite of the fact that Group P-70 had 70 continuously rewarded trials just prior to extinction.



FIG. 2. Response times during extinction.

## DISCUSSION

The present results indicate that the partial reinforcement effect (PRE) of increased resistance to extinction is relatively unaffected by continuous rein-

forcement following the partial reinforcement. After the present experiment was completed, a study by Quartermain and Vaughan (1961) was published which bears on the same issue. Using lever pressing in a Skinner box, they compared resistance to extinction of rats given acquisition training with either 800 responses of which a random 10% were reinforced, or 400 responses at 10% reinforcement followed by 400 continuously reinforced responses. Their results were entirely consistent with those reported here. These findings demonstrate that discriminability of the transition from training to extinction cannot be taken as explaining much of the partial reinforcement effect.

Although it could be assumed that discrimination has some small effect since in the present study Group P-70 extinguished slightly faster than Group P-0, the fact that there was no difference in resistance to extinction between Groups P-0 and P-25 presents a problem for the discrimination hypothesis. It might be argued that 25 trials were not sufficient for Ss to have noticed a change in the reinforcement schedule, and, hence, a difference between Groups P-0 and P-25 should not have been expected. However, the schedule change clearly influenced running performance. The response speeds of Group P-25 and P-70 Ss increased markedly during the 25 continuously reinforced trials, eventually equaling the speeds of the Group C-70 Ss. This increase in performance can be taken as evidence that the Ss "noticed" the change.

Another proposal might be that Group P-25's resistance to extinction could have been increased by the fact that it had 25 more trials than Group P-0. However, Group C-70 had 70 more trials than did Group C-0, and these C groups did not differ in resistance to extinction. Thus, with the degree of training given in the present experiment, additional trials had no effect on resistance to extinction.

The present experiment can be added

to a growing amount of empirical data indicating that the discrimination hypothesis cannot adequately account for the PRE. Marx (1958) has shown that if certain significant features of the goal situation (i.e., the food cup) are omitted on half of the extinction trials, rate of extinction is decreased rather than increased as would be predicted from a discrimination analysis. The discrimination theory has been most successful in accounting for the fact that rate of extinction usually varies directly with the percentage of reward when number of trials or number of reinforcements are held constant. However, using a design which pitted percentage of reinforcement against number of nonreinforcements, Lawrence, Festinger, and Theios (summarized by Festinger, 1961) found that resistance to extinction was independent of the reinforcement percentage, but was determined primarily by the *number* of nonreinforced trials Ss experienced during training.

The present data as well as these other studies suggest that the discrimination hypothesis is inadequate as an interpretation of partial reinforcement. The present data indicate that an adequate theory of partial reinforcement must have constructs representing relatively permanent effects of nonreinforcement which can be sustained through blocks of continuous reinforcement. At least four existing theories of partial reinforcement employ constructs of this character. These theories include Amsel (1958), Estes (1959), Logan (1960), and Festinger (1961). The present results do not enable one to discriminate among these theories.

## SUMMARY

Three groups of 14 rats each received 70 trials of random 40% reinforcement in a runway. Then different numbers of continuously reinforced trials were interpolated between the 40% reinforcement and extinction. Group P-0 received none, Group P-25 received 25, and Group P-70 received 70 interpolated 100% reinforced trials. Two control groups (N = 9 each) received only continuous rein-

forcement during training: Group C-0 was
extinguished with partial Group P-0; Group
C-70 was given 70 additional reinforced trials
and extinguished with Group P-70.

There was no significant difference in
resistance to extinction between the two
continuously reinforced groups. Group P-70
was very significantly $(P < .001)$ more re-
sistant to extinction than the two continuously
reinforced groups. There was no significant
difference in resistance to extinction between
Groups P-0 and P-25, but both of these groups
were slightly more resistant than Group P-70.
These results question the adequacy of the
discrimination hypothesis that the partial
reinforcement extinction effect results from
difficulty in discriminating the transition
from training to extinction.

## REFERENCES

AMSEL, A. The role of frustrative nonreward
in noncontinuous reward situation. *Psy-
chol. Bull.*, 1958, 55, 102–119.
BITTERMAN, M. E., FEDDERSON, W. E.,
& TYLER, D. W. Secondary reinforcement
and the discrimination hypothesis. *Amer.
J. Psychol.*, 1953, 66, 456–464.
ESTES, W. K. The statistical approach to
learning theory. In S. Koch (Ed.),
*Psychology: A study of a science.* Vol. 2.
New York: McGraw-Hill, 1959.
FESTINGER, L. The psychological effects
of insufficient reward. *Amer. Psychologist*,
1961, 16, 1–11.
LOGAN, F. A. *Incentive.* New Haven: Yale
Univer. Press, 1960.
MARX, M. H. Resistance to extinction as a
function of continuous or intermittent
presentation of a training cue. *J. exp.
Psychol.*, 1958, 56, 251–255.
MOWRER, O. H., & JONES, H. M. Habit
strength as a function of the pattern of
reinforcement. *J. exp. Psychol.*, 1945,
43, 293–311.
QUARTERMAIN, D., & VAUGHAN, G. M.
Effect of interpolating continuous rein-
forcement between partial training and
extinction. *Psychol. Rep.*, 1961, 8, 235–237.

# THE EFFECTS OF LANGUAGE HABITS ON THE ACQUISITION AND RETENTION OF VERBAL ASSOCIATIONS [1]

LEO POSTMAN

*University of California*

The experiments reported in this paper investigate the effects of unit-sequence habits on the acquisition and retention of verbal materials. By unit-sequence habits we refer to associative connections between verbal items established through linguistic usage (Underwood & Postman, 1960). The transfer effects of such language habits may be either positive or negative, depending on the relationship between $S$'s pre-experimental associations and those prescribed in the learning task.

The amount of transfer from pre-experimental habits should increase with the frequency of usage of the verbal units in the language. The higher the frequency of usage, the larger on the average is the number of different contexts in which an item is likely to appear and hence the larger the number of different associations which it acquires. Conventional indices of meaningfulness such as association value and $m$ value do in fact correlate highly with frequency of usage (Postman, 1961; Underwood & Schulz, 1960). Both unit-sequence facilitation and unit-sequence interference may then be expected to increase as a function of frequency of usage. When the sequence to be learned involves high-frequency items, at least some of the prescribed associations are likely to agree with pre-experimental habits or can be readily established through direct mediational links. However, the number of pre-experimental associations which can compete with the prescribed ones will also increase as a function

of frequency of usage. With low-frequency materials, the prescribed sequences will have little pre-experimental strength and few direct mediators will be available, but there will also be few pre-experimental associations to produce interference.

The conditions of facilitation and interference should influence recall as well as acquisition. When there is unit-sequence facilitation, it is assumed that the strength of the pre-experimental habits has generalized to the prescribed associations within the list. The greater the generalized habit strength, the more resistant the items in the experimental list will be to interference. When there is unit-sequence interference, it is assumed that competing pre-experimental habits have been unlearned during acquisition and have recovered over time to become sources of interference at recall.

An earlier study (Postman, 1961) investigated the acquisition and retention of serial lists of words of high and low frequency of usage (Lists HF and LF). List HF was learned faster than List LF. At the same time, however, the rate of misplaced responses was substantially greater for List HF. The positive relationship between speed of learning and error rate points to covariation of unit-sequence facilitation and interference as a function of word frequency. There were no significant differences in the amount of retention for the two kinds of materials. However, misplaced responses at recall recovered at a faster rate for List HF than for List LF. While there was no evidence for superior retention of materials of high meaningfulness, the original expectation of an inverse relationship between word frequency and recall was not borne out. This expectation had been based on the

assumption that competition from extra-experimental associations would increase directly with word frequency. The difference in recovery of errors is in accord with this assumption. However, the fact that the amounts recalled were comparable in spite of a substantial difference in the incidence of overt errors suggests that increases in unit-sequence interference may be offset in part by unit-sequence facilitation.

The present experiments extend the analysis of the conditions of unit-sequence facilitation and interference. A first step was to vary the word frequency of stimuli (Ss) and responses (Rs) in a study investigating the acquisition and retention of paired associates (Exp. I). There is considerable evidence that speed of learning varies directly with the meaningfulness of Ss and Rs, and that this relationship is considerably more pronounced and consistent for Rs than for Ss (Cieutat, Stockwell, & Noble, 1958; Hunt, 1959; Kimble & Dufort, 1955; L'Abate, 1959; Mandler & Campbell, 1957; Morikawa, 1959; Underwood & Schulz, 1960). In these experiments the units scaled for meaningfulness (e.g., trigrams or paralogs) were either nonsense items or, more usually, included both nonsense items and words. Letter-sequence habits as well as unit-sequence habits contribute to the differences in speed of learning obtained with such materials (Underwood & Schulz, 1960). The lists used in Exp. I consisted entirely of words so that variations in speed of learning and amount of retention could be attributed primarily to the influence of unit-sequence habits. It was expected that both unit-sequence facilitation and unit-sequence interference would increase with word frequency, with the relative amounts of interference growing more rapidly as a function of S frequency than of R frequency. To the extent that paired items are practiced in the order of their presentation, competing associations elicited by the Ss should be more effective in delaying acquisition than those evoked by the Rs. These considerations suggest that speed of learning should rise to a maximum

and then decline as the word frequency of Ss is increased. On the other hand, speed of learning should vary directly with the word frequency of Rs.

No prior studies of the effects of S and R meaningfulness on the long-term retention of paired associates are available. If the associative context of the Ss rather than the Rs is critical in determining the amount of effective competition at recall, differences in the amount of retention should be determined largely by the degrees of unit-sequence facilitation and interference falling on the S terms. Thus, the differences among S conditions in learning should predict recall better than those among the R conditions.

According to the present analysis, a sample of high-frequency Ss will activate strong pre-experimental habits whereas a sample of low-frequency Ss will not. It follows that manipulation of the conditions of facilitation and interference within a list should be more effective with low-frequency Ss than with high-frequency Ss. When pre-experimental associative probabilities are low, the introduction of otherwise unlikely sources of facilitation and interference based on language habits should produce major effects on learning and retention. On the other hand, when the associative probabilities are high, the introduction of known associates or competitors should add relatively little to the amounts of facilitation and interference from uncontrolled sources. This prediction was tested in Exp. II in which the pre-experimental probability of the prescribed associations was varied for Ss of high and low word frequency.

Experiment III was designed to provide a further test of the assumption that the higher the word frequency of a series of items, the greater is the amount of overlap among their associative contexts. The method of verbal discrimination learning was used. To the extent that discriminability depends on the availability of differential responses to the members of a pair, speed of discrimination learning should increase with the meaningfulness of the paired items.

## TABLE 1

### STIMULUS AND RESPONSE TERMS IN EXP. I

| $S_h$ | $R_h$ | $S_m$ | $R_m$ | $S_l$ | $R_l$ |
|---|---|---|---|---|---|
| BUILDING | ANSWER | ARBOR | BASIN | BRAMBLE | ABBESS |
| COUNTRY | COLOR | BURLAP | DOGMA | CAUCUS | BUFFOON |
| DOCTOR | DINNER | CINDER | FETISH | DECOY | DOTAGE |
| GARDEN | FIGURE | DISCORD | HERMIT | FARTHING | HAREM |
| HUSBAND | MOMENT | GROCER | MINSTREL | GULLET | OBOE |
| LETTER | PROBLEM | LOTION | OMEN | LORRY | PREFIX |
| MORNING | REASON | MAGNATE | RELIC | OXIDE | RAMROD |
| PAPER | SHOULDER | OATMEAL | SUFFRAGE | PESTLE | STANZA |
| STORY | TABLE | TRAITOR | TEMPEST | SEQUEL | TENURE |
| WINDOW | WOMAN | WHISKER | WAFER | WICKET | WAMPUM |

Such a relationship was found with pairs of nonsense syllables by Runquist and Freeman (1960). However, as the associates elicited by the items increase in number, they should begin to lose distinctiveness because of the increasing probability of overlap in associative context. Acquisition will then require the unlearning of differential responses which produce errors of generalization, but such responses may be expected to recover with the passage of time. Thus, speed of acquisition and amount of retention for the verbal discrimination task should first increase and then decrease as a function of word frequency.

## EXPERIMENT I

### Method

*Materials.*—The learning materials were two-syllable nouns of high, medium, and low frequency of usage. Each frequency range was represented by two lists of 10 words each, one of which served as the S list and the other as the R list ($S_h$, $S_m$, $S_l$; $R_h$, $R_m$, $R_l$). These materials were used in the construction of nine paired-associate lists representing all the possible combinations of word frequencies of the S and R items. The lists of words are shown in Table 1.

The lists sample three frequency ranges in the "L" count of Thorndike and Lorge (1944). The numbers of occurrences in 4.5 million are 1000–3300, 10–33, and 1–3 for words of high, medium, and low frequency, respectively. The mean rated familiarity and *m* values, which were obtained in prior standardization procedures, are correlated with frequency of usage. The words were chosen so as to minimize pre-experimental associations between items in the S lists and items in the R lists. Norms of free association obtained from 1000 students at the University of California were used for this purpose.

There was no duplication of first letters in any of the S or R lists. For all nine combinations of word frequencies the S list and the R list had six first letters in common. Four different pairings of Ss and Rs were used equally often with each of the nine lists.

*Procedure.*—The lists were presented on a Hull-type memory drum at a 2:2 rate (S alone for 2 sec. and S and R together for 2 sec.), with an 8-sec. intertrial interval. Learning was to a criterion of one perfect trial. Four different orders of presentation were used to minimize serial learning. Each of the four random orders was used as a starting order equally often.

Retention was tested by relearning either 30 sec. or 7 days after the end of original practice. Relearning was for five trials or to criterion, whichever took the longer.

*Subjects.*—With nine types of lists and two retention intervals, there were 18 groups of 16 Ss each. All Ss were undergraduate students who had English as their native language and were naive to rote-learning experiments. The Ss were assigned to conditions in blocks of 18, with 1 S per block for each combination of lists and retention intervals. The running order within each block was determined by a table of random numbers, as was the assignment to different pairings and starting orders. No Ss were lost because of failure to learn.

### Results

*Original learning.*—There were no significant differences between the

## TABLE 2
MEAN NUMBERS OF TRIALS TO CRITERION: EXP. I

| R Terms | Stimulus Terms | | | | | | Over-All Mean |
|---|---|---|---|---|---|---|---|
| | $S_h$ | | $S_m$ | | $S_l$ | | |
| | Mean | SD | Mean | SD | Mean | SD | |
| $R_h$ | 10.5 | 6.3 | 10.5 | 40. | 12.1 | 6.9 | 11.0 |
| $R_m$ | 12.9 | 6.0 | 12.2 | 54. | 15.7 | 8.8 | 13.6 |
| $R_l$ | 16.2 | 8.5 | 11.2 | 4.0 | 16.3 | 6.5 | 14.6 |
| Over-all Mean | 13.2 | | 11.3 | | 14.7 | | |

30-sec. and 7-day groups learning a given type of list. The results of the two groups learning the same materials have, therefore, been combined. Table 2 shows the mean trials to criterion for each of the nine kinds of lists and also summarizes the average trends produced by the word frequency of the Ss and Rs. Speed of learning first increases and then decreases as a function of the word frequency of Ss but varies directly with the word frequency of Rs. The over-all effects of the word frequency of both Ss and Rs are significant beyond the .01 level ($F = 6.15$ for Ss and 7.24 for Rs, $df = 2/279$ for both). The S $\times$ R interaction is not significant ($F = 1.54$, $df = 4/279$).

*Errors during learning.*—Table 3 shows the mean percentages of misplaced Rs during learning. Such responses account for the great majority of overt errors. The percentages are based on opportunities (total number of presentations minus number of correct responses) and thus are independent of speed of learning.

For both Ss and Rs the mean percentages of errors are clearly larger when word frequency is high than when it is medium or low. There are only small differences between the latter two conditions. Following arc-sine transformation and with $2/279$ $df$, the $F$ ratio for Ss is 12.44 ($P < .001$) and 3.11 for Rs ($.02 < P < .05$). There is also a significant S $\times$ R interaction ($F = 2.62$, $df = 4/279$, $.02 < P < .05$). When the word frequency of either the Ss or the Rs is high, the percentage of errors remains uniformly large, regardless of the value of the other variable.

Table 3 also shows for each type of list the product-moment correlation between trials to criterion and percentage of misplaced Rs during learn-

## TABLE 3
MEAN PERCENTAGES OF MISPLACED RESPONSES AND CORRELATION BETWEEN PERCENTAGES AND TRIALS TO CRITERION: EXP. I

| R Terms | Stimulus Terms | | | | | | | | | Over-All Mean |
|---|---|---|---|---|---|---|---|---|---|---|
| | $S_h$ | | | $S_m$ | | | $S_l$ | | | |
| | Mean | SD | r | Mean | SD | r | Mean | SD | r | |
| $R_h$ | 26.6 | 17.5 | .61* | 23.0 | 17.9 | .55* | 26.3 | 17.6 | .55* | 25.3 |
| $R_m$ | 27.1 | 12.9 | .26 | 16.4 | 10.1 | .59* | 16.9 | 12.1 | .30 | 20.1 |
| $R_l$ | 30.0 | 15.3 | .34 | 14.8 | 12.1 | .61* | 14.7 | 7.6 | .26 | 19.8 |
| Over-all Mean | 27.9 | | | 18.1 | | | 19.3 | | | |

* $P < .01$.

TABLE 4

MEAN NUMBERS OF ITEMS RECALLED IN EXP. I

| R Terms | 30-Sec. Recall | | | | | | | 7-Day Recall | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stimulus Terms | | | | | | Over-All Mean | Stimulus Terms | | | | | | Over-All Mean |
| | $S_h$ | | $S_m$ | | $S_l$ | | | $S_h$ | | $S_m$ | | $S_l$ | | |
| | Mean | SD | Mean | SD | Mean | SD | | Mean | SD | Mean | SD | Mean | SD | |
| $R_h$ | 9.4 | .7 | 9.2 | .7 | 9.2 | .8 | 9.3 | 5.3 | 1.7 | 5.4 | 2.1 | 5.2 | 1.9 | 5.3 |
| $R_m$ | 9.2 | .7 | 8.8 | .8 | 9.0 | 1.2 | 9.0 | 4.9 | 2.0 | 5.1 | 1.9 | 3.8 | 2.1 | 4.6 |
| $R_l$ | 8.7 | .9 | 9.1 | 1.2 | 8.6 | 1.0 | 8.8 | 4.7 | 2.3 | 5.2 | 1.9 | 3.5 | 2.4 | 4.5 |
| Over-all Mean | 9.1 | | 9.0 | | 8.9 | | | 5.0 | | 5.2 | | 4.2 | | |

ing. All the correlations are positive, i.e., the higher the error rate, the more slowly is criterion reached. This relationship is especially pronounced and statistically significant for Cond. $S_m$ and $R_h$.

The inverse relationship between error rate and speed of learning found for individuals within groups clearly does not hold for materials. On the R side, both speed of learning and error rate are maximal for Cond. $R_h$; on the S side, error rate is maximal, and speed of learning is intermediate, for Cond. $S_h$. Within a given group, however, rate of acquisition varies with Ss' susceptibility to interference, especially when learning is fast as in Cond. $S_m$ and $R_h$. When correct Rs are acquired rapidly, a low percentage of misplaced Rs is favorable to attainment of the criterion, i.e., failures to respond are eliminated more readily than overt errors. When correct Rs increase relatively slowly, the criterion score is less sensitive to the type of error.

*Recall.*—Table 4 shows the mean numbers of Rs recalled by the various groups. The drops on the 30-sec. test vary inversely with speed of learning although the correlation is not perfect. The amounts of reten-

tion loss (mean 30-sec. recall − mean 7-day recall) as a function of the word frequency of Ss and Rs are shown in Fig. 1. The losses during the 7-day interval are smallest for $S_m$ and greatest for $S_l$, with $S_h$ yielding an intermediate amount of forgetting. The
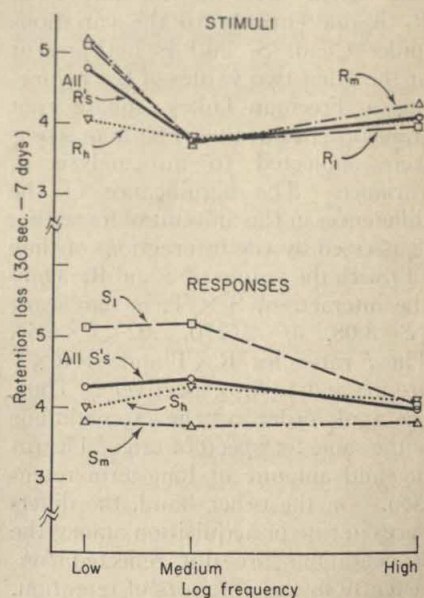


FIG. 1. Amounts of retention loss (mean 30-sec. recall − mean 7-day recall) as a function of the word frequency of Ss and Rs (Exp. I).

## TABLE 5
### Numbers of Misplaced Responses at Recall: Exp. I

| R Terms | 30-Sec. Recall | | | | | | | | 7-Day Recall | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stimulus Terms | | | | | | Over-All | | Stimulus Terms | | | | | | Over-All | |
| | $S_h$ | | $S_m$ | | $S_l$ | | | | $S_h$ | | $S_m$ | | $S_l$ | | | |
| | Total | N | Total | N | Total | N | Total | N | Total | N | Total | N | Total | N | Total | N |
| $R_h$ | 5 | 4 | 8 | 7 | 5 | 5 | 18 | 16 | 17 | 8 | 12 | 7 | 13 | 9 | 42 | 24 |
| $R_m$ | 5 | 5 | 9 | 7 | 4 | 4 | 18 | 16 | 9 | 6 | 4 | 4 | 9 | 8 | 22 | 18 |
| $R_l$ | 11 | 9 | 2 | 1 | 6 | 5 | 19 | 15 | 9 | 6 | 4 | 2 | 10 | 6 | 23 | 14 |
| Over-all | 21 | 18 | 19 | 15 | 15 | 14 | | | 35 | 20 | 20 | 13 | 32 | 23 | | |

Note.—N = Number of Ss giving misplaced responses.

average difference between $S_h$ and $S_m$ is, however, clearly smaller than that between $S_l$ and $S_m$. The trends for the three values of R are comparable. There are no consistent variations in retention loss as a function of the word frequency of Rs. The slight over-all advantage of $R_h$ is due entirely to the variations under Cond. $S_l$ and is not present at the other two values of S. Following a Freeman-Tukey square root transformation, the retention scores were subjected to an analysis of variance. The significance of the differences in the amount of forgetting is assessed by the interactions of time (T) with the values of S and R. Only the interaction, S × T, is significant ($F = 3.08$, $df = 2/270$, $.02 < P < .05$). The $F$ ratios for R×T and S×R×T are .88 and 1.08, respectively. Thus, the rank order of the S conditions is the same for speed of original learning and amount of long-term retention. On the other hand, the differences in rate of acquisition among the R conditions are not reflected consistently in the amounts of retention.

*Misplaced responses at recall.*—The numbers of misplaced Rs at recall and the numbers of Ss contributing them are shown in Table 5. Such errors account for only a small fraction of the total retention losses. A large proportion of Ss—99/144 in the 30-sec. groups and 88/144 in the 7-day groups—failed to give misplaced Rs. The observed frequencies indicate a higher rate of recovery of intralist intrusions for $S_h$ and $S_l$ than for $S_m$, and also a higher rate for $R_h$ than for $R_m$ and $R_l$. In view of the high proportion of zero scores, the differences in temporal trends were not evaluated statistically and must be interpreted with caution. It is interesting to note, however, that the apparent differences in rate of recovery are related to the amount of forgetting for the S conditions but not the R conditions. The low rate of intralist intrusions suggests that Ss had considerable success in rejecting incorrect Rs.

*Relearning.*—The mean numbers of trials to criterion in relearning are shown in Table 6. The relative amounts of retention loss correspond to those found at recall: the increase in trials to criterion is greater for $S_h$ and $S_l$ than for $S_m$ whereas the word frequency of Rs has little effect. The interaction, S × T, is significant ($F = 4.52$, $df = 2/270$, $.02 < P < .05$).

TABLE 6

MEAN NUMBERS OF TRIALS IN RELEARNING TO CRITERION: EXP. I

| R Terms | 30-Sec. Recall | | | | | | | 7-Day Recall | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stimulus Terms | | | | | | Over-All Mean | Stimulus Terms | | | | | | Over-All Mean |
| | $S_h$ | | $S_m$ | | $S_l$ | | | $S_h$ | | $S_m$ | | $S_l$ | | |
| | Mean | SD | Mean | SD | Mean | SD | | Mean | SD | Mean | SD | Mean | SD | |
| $R_h$ | 2.2 | 1.7 | 2.6 | 2.2 | 2.1 | 1.2 | 2.3 | 5.1 | 1.8 | 3.8 | 1.8 | 4.1 | 1.9 | 4.3 |
| $R_m$ | 2.1 | .9 | 3.1 | 2.2 | 2.4 | 2.1 | 2.5 | 4.4 | 2.5 | 3.5 | 1.4 | 5.9 | 2.8 | 4.6 |
| $R_l$ | 2.5 | 1.1 | 1.9 | 1.0 | 2.8 | 1.4 | 2.4 | 5.2 | 2.5 | 3.2 | 1.2 | 5.1 | 4.3 | 4.5 |
| Over-all Mean | 2.3 | | 2.5 | | 2.4 | | | 4.9 | | 3.5 | | 5.0 | | |

The $F$ ratios for $R \times T$ and $S \times R \times T$ are .01 and .85, respectively.

## EXPERIMENT II

### Method

*Materials.*—There were two lists of 10 paired associates, one with S terms of high word frequency (List $S_h$), and the other with S terms of low word frequency (List $S_l$). The S terms in both lists were two-syllable nouns drawn from the same pools of items as the words in Exp. I. Half the Rs in each list were highly probable associates ($A_h$) of the S words, i.e., they were the primary Rs to these words in free association norms ob-

tained from 1000 students at the University of California. The other half of the Rs were associates of low probability ($A_l$), having been given by only 2/1000 Ss in the normative group. Thus, there were four types of pairs: $S_h$-$A_h$ and $S_h$-$A_l$ in one list, $S_l$-$A_l$ and $S_l$-$A_l$ in the other. All the Rs were nouns of high frequency of usage falling within the AA and A ranges of the General Count of Thorndike and Lorge (1944). Within each of the four classes of pairs three of the Rs were two-syllable words and two Rs were one-syllable words.

There was no duplication of first letters in either of the S lists. One first letter was duplicated in each of the two R lists. In both

TABLE 7

STIMULUS-RESPONSE PAIRS USED IN EXP. II

| Stimulus | $A_h$ | Assoc. Prob.[a] | Stimulus | $A_l$ | Assoc. Prob.[a] |
|---|---|---|---|---|---|
| List $S_h$ | | | | | |
| GARDEN | FLOWER | 592 | BUILDING | SIZE | 2 |
| MOMENT | TIME | 255 | COUNTRY | VALLEY | 2 |
| ORDER | COMMAND | 115 | LETTER | PERSON | 2 |
| TROUBLE | PROBLEM | 132 | REASON | NEED | 2 |
| WINDOW | GLASS | 252 | SHOULDER | BURDEN | 2 |
| List $S_l$ | | | | | |
| CAUCUS | MEETING | 156 | BRAMBLE | PAIN | 2 |
| GULLET | THROAT | 180 | DECOY | ARMY | 2 |
| HAREM | WOMAN | 340 | OXIDE | COLOR | 2 |
| RAMROD | GUN | 258 | TENURE | LIFE | 2 |
| STANZA | MUSIC | 282 | WAMPUM | DOLLAR | 2 |

[a] Number of Ss out of 1000 giving response on test of free association.

## TABLE 8

MEAN NUMBERS OF TRIALS TO CRITERION OF 5/5 FOR EACH CLASS OF PAIRS AND
10/10 FOR TOTAL LISTS IN ORIGINAL LEARNING AND RELEARNING: EXP. II

| Pairs | 30-Sec. Groups Stimulus List | | | | 7-Day Groups Stimulus List | | | |
|---|---|---|---|---|---|---|---|---|
| | $S_h$ | | $S_l$ | | $S_h$ | | $S_l$ | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| | | | | Original Learning | | | | |
| $A_h$ | 4.6 | 2.7 | 1.9 | .9 | 5.4 | 3.7 | 1.6 | .5 |
| $A_l$ | 4.3 | 2.0 | 2.8 | 1.0 | 4.2 | 1.5 | 2.6 | 1.5 |
| Total list | 5.4 | 2.9 | 2.9 | 1.0 | 6.2 | 3.4 | 2.8 | 1.5 |
| | | | | Relearning | | | | |
| $A_h$ | 1.4 | .8 | 1.0 | 0 | 3.3 | 1.9 | 1.7 | .6 |
| $A_l$ | 1.8 | 1.0 | 1.2 | .6 | 2.6 | 1.3 | 2.2 | .6 |
| Total list | 1.9 | 1.2 | 1.2 | .6 | 4.7 | 1.9 | 2.5 | .8 |

lists the S and R terms had six first letters in common. Within a list none of the Rs appeared in the association tables for any of the stimuli except the one with which it was paired. Table 7 presents the lists and for each pair shows the frequency of the R term in the association norms.

*Procedure.*—The lists were presented on a Hull-type memory drum at a 2:2 rate with an 8-sec. intertrial interval. Learning was to one perfect trial. There were four different orders of presentation each of which was used as a starting order equally often. Retention was tested by relearning either 30 sec. or 7 days after the end of original learning. Relearning was for five trials or to criterion, whichever took the longer.

*Subjects.*—With two lists and two retention intervals, the design comprised four groups of 16 Ss each. All Ss were undergraduate students who had English as their native language and were naive to rote-learning experiments. The Ss were assigned to conditions in blocks of 4, with 1 S per block for each combination of lists and retention intervals. Assignment to conditions within a block was determined by a table of random numbers as was the assignment to different starting orders. No Ss were lost because of failure to learn.

## Results

*Original learning.*—The mean numbers of trials to a criterion of 5/5 for each type of pair as well as to a criterion of 10/10 for the entire list are shown in Table 8. There were no significant differences between the 30-sec. and 7-day groups learning a given kind of list. Pairs $S_l$-$A_h$ were mastered extremely rapidly, with the large majority of Ss reaching the criterion of 5/5 in either one or two trials. Pairs $S_l$-$A_l$ were learned more slowly than Pairs $S_l$-$A_h$. However, with half the pairs learned almost immediately, acquisition of the entire list was very rapid. By contrast, Pairs $S_h$-$A_h$ were not learned faster than Pairs $S_h$-$A_l$; in fact, there was a trend in the opposite direction. The over-all rate of learning was significantly slower for List $S_h$ than for List $S_l$ ($t = 4.97$, $df = 62$, $P < .01$). Analysis of variance of the criterion scores on the four kinds of pairs shows the interaction, Pairs × Lists, to be significant ($F = 7.08$, $df = 1/62$, $P < .01$). Thus, the association norms predict speed of learning for List $S_l$ but not for List $S_h$.

*Errors during learning.*—Misplaced Rs were much more common during

the acquisition of List $S_h$ than of List $S_l$. Twenty-two of the $S$s learning List $S_h$ made such errors, and only 6 of the $S$s learning List $S_l$. When the numbers of misplaced Rs are expressed as percentages based on opportunities, the mean percentage for List $S_h$ is 27.6 ($SD = 20.0$; Median $= 26.8$). For List $S_l$ the mean of the extremely skewed distribution is 6.4 ($SD = 16.1$; Median $< 1.0$). The difference between the two distributions is significant beyond the .01 level by Wilcoxon's test for signed ranks, with $S$s in the same block treated as paired replicates.

A total of 23 importations from outside the list was contributed by 13 of the $S$s learning List $S_l$; 3 of the $S$s learning list $S_h$ contributed a total of five such errors. The difference between the proportions of $S$s making such errors is significant beyond the .01 level ($\chi^2 = 6.75$; $df = 1$). All of the importations appeared in the association tables for the appropriate stimuli. Of 23 importations into List $S_l$, 19 were substitutions of primary associates for prescribed $A_l$ responses; no such cases occurred during the learning of List $S_h$.

*Recall.*—The recall scores for the four kinds of pairs are shown in Table 9. When we consider the two types of pairs within each list, we find clearly superior retention of the high-probability pairs in List $S_l$ but no difference in List $S_h$. Comparison of the corresponding pairs in the two lists shows retention to be higher for Pairs $S_l$-$A_h$ than for Pairs $S_h$-$A_h$. By contrast retention is poorer for Pairs $S_l$-$A_l$ than for Pairs $S_h$-$A_l$.

In view of the virtual absence of any variation in the recall scores on the 30-sec. test, the following procedure was used to evaluate the statistical significance of the differences in the amount of forgetting.

Within each block, the difference between the scores obtained by $S$s in the 30-sec. and 7-day groups was determined for each type of pair, and the distribution of differences was subjected to an analysis of variance. The over-all difference between Lists is not significant ($F < 1$), whereas the interaction of Lists with Pairs is significant beyond the .01 level ($F = 8.59$, $df = 1/30$). The relationship between the associative rank of the Rs and the rate of forgetting varies reliably with the word frequency of the Ss.

*Errors at recall.*—The numbers of misplaced Rs at recall were small and did not increase during the retention interval. The few intralist errors that were made occurred almost exclusively during the recall of List $S_h$, at the rate of approximately .5 per $S$ on both tests. By contrast, importations from outside the list showed a systematic temporal trend. On the 30-sec. test there were only 2 $S$s giving such Rs, both on Pairs $S_l$-$A_l$. The frequencies of importations on the 7-day test and the numbers of $S$s ($N$) contributing them were as follows: $S_h$-$A_h$—4 ($N = 4$); $S_h$-$A_l$— 5 ($N = 5$); $S_l$-$A_h$—5 ($N = 5$); $S_l$-$A_l$— 19 ($N = 13$). When the 30-sec. and 7-day $S$s assigned to the same block are compared, the number of cases in which there is an increase in

TABLE 9

MEAN NUMBERS OF PAIRS RECALLED IN
EXP II

| Pairs | 30-Sec. Recall Stimulus List | | | | 7-Day Recall Stimulus List | | | |
|---|---|---|---|---|---|---|---|---|
| | $S_h$ | | $S_l$ | | $S_h$ | | $S_l$ | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| $A_h$ | 4.8 | .4 | 5.0 | 0 | 3.2 | 1.2 | 4.1 | .8 |
| $A_l$ | 4.5 | .5 | 4.8 | .4 | 3.0 | 1.3 | 2.2 | 1.2 |
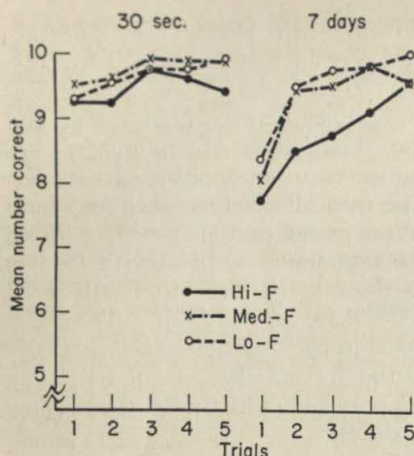
Fig. 2. Mean numbers of correct responses in relearning of verbal discrimination (Exp. III).

importations during the retention interval is significantly greater for List $S_l$ than for List $S_h$ ($\chi^2 = 4.52$, $df = 1$, $.02 < P < .05$). It is clear that this difference is primarily a function of the number of imported substitutions for low-probability associates. In the recall of List $S_h$, none of these substitutions were primary associates of the appropriate $S$s; in the case of List $S_l$, 16/19 substitutions were primary associates.

*Relearning.*—The mean numbers of trials to relearn to a criterion of 5/5 on each type of pair and 10/10 for the entire list are shown in Table 8. The differences in the amount of forgetting as measured by speed of relearning correspond to those found at recall. The procedure used in analyzing the temporal trends in recall was also applied to the relearning scores. The over-all difference between Lists was again not reliable ($F = 1.53$), whereas the interaction of Lists with Pairs is significant ($F = 5.11$, $df = 1/30$, $.02 < P < .05$).

## EXPERIMENT III

### Method

*Materials.*—The learning materials were three lists of 10 pairs of words each. One list (List HF) was composed of words of high frequency of usage, one list (List MF) of words of medium frequency, and one list (List LF) of words of low frequency. The words used in the construction of the lists were the same as those in the S lists and R lists of Exp. I (see Table 1). Two different pairings of the words in each list were used equally often. For purposes of verbal discrimination learning, one of the items in each pair was designated as correct. Each item in a pair was correct for half the $S$s.

*Procedure.*—The members of a pair were presented one above the other in the window of a Hull-type memory drum. The rate of presentation was 1.5 sec. for the pair alone, and 1.5 sec. for the pair and the correct R together. The intertrial interval was 3 sec. Four different orders of presentation were used to minimize serial learning. In each order half of the correct words appeared in the upper position and half in the lower position. The position of the correct item in each pair varied from trial to trial. Each of the four orders was used as a starting order equally often. The $S$s began anticipating the correct R on the first trial. Learning was to a criterion of one perfect trial.

Retention was tested by relearning either 30 sec. or 7 days after the end of practice. Relearning was for five trials or to criterion, whichever took the longer.

*Subjects.*—With three types of lists and two retention intervals, there were six groups of 16 $S$s each. All $S$s were undergraduates who had English as their native language; they were not necessarily naive to rote-learning experiments but had had no prior experience with verbal discrimination learning. The $S$s were assigned to conditions in blocks of 6, with 1 $S$ per block for each combination of lists and retention intervals. The running order within each block and the assignment to different starting orders were determined by tables of random numbers. No $S$s were lost because of failure to learn.

### Results

*Original learning.*—There were no significant differences between the 30-sec. and 7-day groups learning a given kind of list. The mean numbers of trials to criterion for the com-

bined groups were as follows: HF—6.4 ($SD = 4.0$); MF—4.5 ($SD = 1.5$); LF—5.3 ($SD = 2.5$). The effects of word frequency on acquisition parallel those obtained on the S side in paired-associate learning in Exp. I, i.e., List MF is learned faster than Lists HF and LF. The variation in criterion scores is significant ($F = 3.63$, $df = 2/93$, $.02 < P < .05$).

*Recall and relearning.*—Figure 2 shows the mean numbers of correct Rs on the five trials of relearning 30 sec. and 7 days after the end of practice. The 30-sec. test reflects the rank order of the conditions in original learning. Immediate relearning proceeds less steadily for List HF than for the other two lists. The disadvantage of List HF becomes pronounced on the 7-day test. A noteworthy fact is that the difference between List HF and the other two lists increases between Trial 1 and Trial 2 and that this separation is maintained through Trial 4. The conditions which slowed up the acquisition of List HF rapidly come into play again in relearning. The separation between Lists MF and LF is consistently small.

The differences on the recall trial are not significant. For the five trials of relearning, the over-all variation among lists is significant ($F = 7.52$, $df = 2/90$, $P < .01$). With performance on List HF poorest on both tests, the Time × Lists interaction falls short of significance ($F = 2.12$, $df = 2/90$). The pattern of differences and the results of the statistical tests are the same when trials to relearn to criterion are considered.

## DISCUSSION

The results of the experiments are consistent with the assumption that unit-sequence interference and unit-sequence facilitation covary as a function of word frequency, with the relative amount of effective interference increasing more rapidly on the S than on the R side. Such systematic changes in the pattern of transfer effects are indicated by the findings on acquisition in Exp. I: (a) rate of overt errors is a positively accelerated function of the word frequency of both Ss and Rs; (b) speed of learning increases steadily with the word frequency of Rs; (c) speed of learning first increases and then decreases as a function of the word frequency of Ss.

Speed of learning is influenced by the word frequency of both Ss and Rs whereas the amount of retention varies significantly only with S conditions. In acquisition the availability of the R units, which is a function of frequency, is a major determinant of the speed of learning (Underwood & Schulz, 1960). Once the Rs are available to the learner, the associates elicited by the Ss are of major importance in determining the speed at which the prescribed associative connections are established. Retention losses occur as incorrect associations which had been unlearned during acquisition recover in strength relative to the prescribed sequences. On the recall test, the competing Rs elicited by the S terms largely determine the amount of effective interference. The rate at which such competing associations recover with the passage of time reflects the balance of positive and negative transfer effects in acquisition. There appears to be no differential decline in the availability of the Rs when well integrated units such as words are used.

Experiment II shows that the pre-experimental probabilities of specific associates have a major influence on learning and retention when the word frequency of the Ss is low but have no differential effects when the word frequency of the Ss is high. In the acquisition of List $S_h$, strong associative connections either existed or were readily established by direct mediational links between any pair of words in the list, and the primary associates had no advantage over associates of low rank. In the case of List $S_l$, on the other hand,

the uncontrolled transfer effects of pre-experimental habits were weak and did not mask the difference between associates of high and low probability.

The results of the retention tests in Exp. II clearly bring out the differences in unit-sequence interference and facilitation falling on the pairs in the two lists. When facilitation by pre-experimental habits obtains for the pairs in both lists ($S_h$-$A_h$ and $S_l$-$A_h$), retention is inversely related to the word frequency of the Ss. Under these conditions the rate of forgetting directly reflects the assumed difference in unit-sequence interference. In the case of the low-probability pairs ($S_h$-$A_l$ and $S_l$-$A_l$) there is interference from language habits in both lists, as indicated by the intralist errors in List $S_h$ and the frequent importation of primary associates into List $S_l$. For these pairs, retention varies directly with word frequency, now reflecting the assumed difference in unit-sequence facilitation. The question arises, of course, whether the differences in recall, and for that matter in acquisition, are simply the result of a set to give primary associates. While such a set was undoubtedly present, the critical finding is that it was effective only with Ss of low word frequency. With Ss of high word frequency, no one associate is clearly dominant, and Rs from outside the list cannot compete effectively with those within the list.

The fact that speed of verbal discrimination learning first increases and then decreases as a function of word frequency (Exp. III) provides additional support for the assumed covariation of unit-sequence facilitation and interference. With homogeneous pairs such as were used in this study, differential effects of R availability are minimized (cf. Runquist & Freeman, 1960), and the relationship parallels that obtained for Ss in paired-associate learning. The availability of strong differential Rs is essential for the achievement of a stable discrimination, but once the number of such Rs exceeds an optimal value, generalization across items counteracts the beneficial effects of R-produced differ-

entiation. It is likely that generalization both within pairs and between pairs contributes to the total effect. While recall was too high to yield reliable differences, List HF is clearly inferior to the other two lists in relearning.

In agreement with the results obtained earlier with serial lists (Postman, 1961), the present experiments offer no support for the assumption that meaningfulness necessarily favors retention. By maximizing either the negative or positive transfer effects of pre-experimental habits, it is possible to obtain either an inverse or a direct relationship between meaningfulness and retention. For an average sample of verbal materials, the best prediction at present is that the differences in retention will be small when degree of original learning is held constant.

## Summary

Three experiments investigated the transfer effects of language habits on the acquisition and retention of verbal associations. They tested the assumption that both positive and negative transfer effects (unit-sequence facilitation and unit-sequence interference) increase as a function of the frequency of usage of words. The balance of unit-sequence interference and facilitation determines the speed of acquisition and the rate at which interferences recover with the passage of time.

In Exp. I lists of paired associates representing all possible combinations of S and R terms of high, medium, and low word frequency were used. Speed of acquisition first increased and then decreased as a function of the word frequency of Ss but varied directly with the word frequency of Rs. Amount of retention did not vary significantly as a function of R conditions whereas the effect of S frequency paralleled that obtained in acquisition. While the availability of Rs and pre-experimental associative probabilities both influence speed of acquisition, amount of retention loss appears to be determined primarily by the recovery of competing associations elicited by the S terms.

In Exp. II associates of high and low pre-experimental probability, as determined by norms of free association, were learned to Ss of high and low word frequency. Pre-experimental associative probability has significant effects on learning and retention when the

word frequency of the Ss is low but not when it is high. Whether or not manipulation of the conditions of facilitation and interference is successful depends on the magnitude of the uncontrolled transfer effects of pre-experimental habits.

The method of verbal discrimination learning was used in Exp. III with pairs of words of either high, medium, or low word frequency. Speed of acquisition first increased and then decreased as a function of the word frequency of the paired items. Thus, the relationship paralleled that obtained for the word frequency of Ss in paired-associate learning. The differences in recall were unreliable, but the high-frequency lists were relearned more slowly than those of medium or low frequency.

## REFERENCES

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, 56, 193–202.

HUNT, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, 57, 262–267.

KIMBLE, G. A., & DUFORT, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, 50, 361–368.

L'ABATE, L. Manifest anxiety and the learning of syllables with different associative values. *Amer. J. Psychol.*, 1959, 72, 107–110.

MANDLER, G., & CAMPBELL, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, 54, 269–273.

MORIKAWA, Y. Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, 43, 437–446.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, 61, 97–110.

RUNQUIST, W. N., & FREEMAN, M. Roles of association value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, 59, 396–401.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73–95.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

# CONDITIONED DIMINUTION OF THE UNCONDITIONED RESPONSE AS A FUNCTION OF THE NUMBER OF REINFORCEMENTS [1]

H. D. KIMMEL           AND           H. S. PENNYPACKER

*University of Florida*                    *Duke University*

Kimble (1961) has recently shown that the amplitude of the unconditioned eyeblink is attenuated by the presence of the CS during reinforced CS-UCS conditioning trials. Assuming that this diminution of the UCR is a manifestation of a conditioned inhibitory process under the control of the CS, Kimble and Ost (1961) varied the CS-UCS interval during conditioning to determine whether or not the degree of inhibition varied as a function of the interstimulus interval. They found that the strength of the conditioned inhibitory process was related to the interstimulus interval in the same way as is the conditioned excitatory process, i.e., the greatest amount of recovery in UCR amplitude occasioned by the omission of the CS occurred when the interval during conditioning had been 0.5 sec.

The purpose of the present study was to extend these eyelid conditioning findings to another classically conditionable human response, the GSR, and to investigate the relationship between this conditioned inhibitory process and another parameter of the conditioning process, the number of reinforcements. It was hypothesized that attenuation of the UCR during reinforcement and recovery of the UCR upon omission of

the CS would both vary positively with variations in the number of reinforcements.

## METHOD

*Subjects.*—Thirty students in introductory psychology at Duke University were Ss. The Ss volunteered to meet a class requirement. They were assigned randomly to three groups of 10 each, to receive different numbers of reinforcements, but were otherwise treated identically.

*Instructions.*—The S read the instructions from a typewritten card. They indicated that this was an experiment on the effect of environmental stimulation upon the GSR and that S's task was to remain relaxed and motionless but to pay attention to the stimuli during the experiment.

*Apparatus.*—The CS was a 1000-cps tone produced by General Radio Company equipment and delivered to S via Trimm Company earphones. The intensity of the tone, rated at the earphones, was 40 db. (re.: .0002 dynes/cm²).

The UCS was an electric shock produced by a Psychological Instruments Company stimulator and delivered to the index and middle fingertips of S's right hand. The shock intensity used during conditioning was 2 ma. at 50 v., for a hypothetical S with a resistance of 25,000 ohms.

The durations of the stimuli were controlled electronically. The CS had a duration of 1.0 sec. and the UCS a duration of 0.1 sec. On paired conditioning trials a delayed paradigm was used with a 0.9-sec. interval between onsets of the two stimuli.

The GSR was measured as the maximum decrease in resistance which occurred within 3 sec. after the offset of a stimulus. The response was picked up from the palm and back of S's left hand by ¾-in. zinc electrodes, covered with a few drops of zinc sulphate solution, in lucite cups filled with saline electrode jelly. It was amplified by a Hunter GSR apparatus and recorded on an Esterline-Angus ink-writing milliammeter with a paper

speed of 3 in/min. All responses were transformed to the square root of conductance change ( $\sqrt{\Delta C}$ ) for statistical purposes.

*Procedure.*—Data were collected in a quiet dark room. An electric fan masked extraneous sounds. The $E$ and the apparatus were located in an adjoining room.

The $S$'s left hand was cleaned with acetone before the GSR electrodes were attached. He was then seated in the experimental room and given the instructions to read before the pickup leads were attached to the GSR electrodes, the shock electrodes were taped to his fingertips, and the earphones were placed on his head.

The $S$ was first given 2 presentations of the CS alone followed by 3 presentations of the shock alone, in intensity increments up to the intensity used during conditioning. This was followed by 4 additional presentations of the CS alone. During the reinforcement period, which occurred next, the three groups of $S$s received either 4, 8, or 16 paired presentations of the CS and UCS. After reinforcement 2 presentations of the shock alone were given and, finally, 4 presentations of the CS alone (i.e., extinction). All presentations (i.e., trials) were separated by intervals of 30 to 60 sec., varied unsystematically by $E$.

## RESULTS

Of primary significance in evaluating the hypotheses was the way in which the amplitude of the UCR varied over trials during reinforcement and the degree to which an increment in amplitude of the UCR was produced by the omission of the CS after reinforcement. Figure 1 shows the average amplitude of GSR for each group of $S$s, as a function of reinforced trials. Also indicated in Fig. 1 are the average amplitudes of GSR on the two test trials of the UCS only (shown in the figure). Each data point represents the mean of 10 $S$s.

Figure 1 indicates that the amplitude of the UCR gradually reduced during reinforcement, although the groups differed in the extent to which this reduction occurred. Statistical analysis of the degree of diminution of GSR amplitude was achieved by
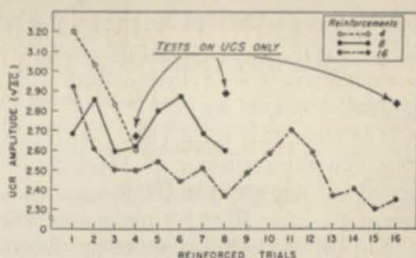


FIG. 1. Average amplitude of the unconditioned GSR during reinforcement and on test trials without the CS.

comparing the average amplitude of GSR on the first two reinforced trials (combined) with the average amplitude of GSR on the last two reinforced trials (combined). Thus, for the group which received four reinforcements the two values used were the average of the first two reinforced trials and the average of the third and fourth reinforced trials. The values used for the group which received eight reinforcements were obtained from the first two reinforced trials and the seventh and eighth reinforced trials, etc. Analysis of variance of these average measures of UCR amplitude during reinforcement showed that the overall reduction in UCR amplitude was statistically significant ( $P < .05$, error variance = .2616) but that it did not interact significantly with the number of reinforcements as had been expected.

Figure 1 also shows the degree to which the omission of the CS produced an increment in UCR amplitude. The exact quantitative nature of the relationship between the number of reinforcements and the degree of increment produced by the omission of the CS is shown in Fig. 2, which presents the average difference (i.e., increment) in GSR amplitude between the last two reinforced trials (combined) and the two test trials of UCS only (combined). A log-

arithmic scale has been used for the abscissa in Fig. 2 to emphasize the linearity of the relationship when expressed in this way.
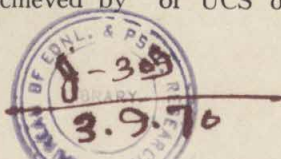
It is clear in Fig. 2 that the degree to which the omission of the CS produced an increment in UCR amplitude increased as a linear function of the logarithm of the number of reinforcements, i.e., as a negatively accelerated growth curve. The analysis of variance of the averaged UCRs on the last two reinforced trials and the two UCS-only trials indicated that the overall increment in UCR amplitude produced by the omission of the CS was statistically significant ($P < .025$) and that the interaction between this increment and the number of reinforcements approached significance ($P < .07$). The linear component of the trend over log N was statistically significant ($P < .025$). Error variance for these tests = .1331.

It was of interest also to examine the relationship between the number of reinforcements and GSR amplitude during extinction. This evaluation was contaminated in part by the fact that all of the $S$s had received two presentations of the UCS alone between the end of reinforcement and the beginning of extinction, which probably tended to obscure any differences in degree of conditioning which might have occurred. For this reason only the last three extinction trials were used to evaluate the degree of conditioning which occurred. When the average amplitude of GSR on these three trials was compared to the average amplitude of the GSR on the last three presentations of the CS alone prior to conditioning it was found that the average amplitude of GSR increased greatly from before to after the reinforcement period and that this increase was largest in the group which received eight reinforcements. The overall difference from before to after reinforcement was highly significant, supporting the conclusion that conditioning occurred, but the interaction between the number of reinforcements and degree of conditioning failed to reach significance.

## DISCUSSION

The results of this experiment support the notion that the CS becomes capable during conditioning of attenuating the amplitude of the UCR. Not only did the amplitude of the UCR diminish during reinforcement, but, when the CS was omitted and the UCS presented alone, the UCR recovered to approximately its original amplitude.

It is of little significance to point out that $S$ was surprised to receive the UCS without the CS and that this surprise, acting as a startle stimulus, may have been responsible for the increment in the GSR on shock-alone trials. The fact that the presence of the CS serves as a warning to $S$ that the shock is coming is, of course, one of the things that are "learned" during the conditioning process. Even when viewed in this way, the reduction-of-surprise function of
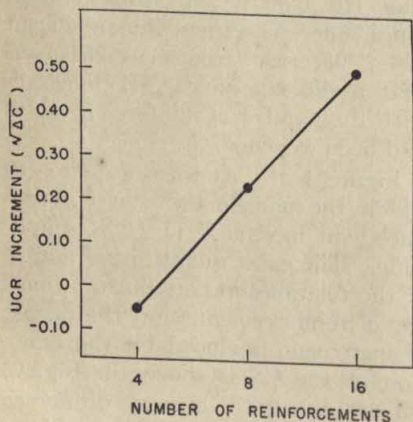


FIG. 2. The relationship between the number of reinforcements and the amount of increment produced by the omission of the CS.

the CS is essentially an inhibitory function. That this inhibitory aspect of the classical conditioning process is highly adaptive to the organism being conditioned is apparent, since a good deal of unnecessary emotional overresponding is brought under control or avoided by the association of the warning signal (i.e., the CS) with the noxious event which is imminent.

The relationship between the present findings and those of an earlier study of disinhibition in GSR conditioning (Kimmel & Fowler, 1961) is of some significance. In the earlier study it was shown that the addition of an extra, or disinhibiting, stimulus to the CS after reinforcement had the effect of producing an increment in the amplitude of the conditioned GSR and that this effect was positively related to the number of reinforcements given prior to the addition of the extra stimulus. The interpretation of the results of the earlier study was based upon the assumption that the CS acquired inhibitory capabilities during reinforcement and that the extra stimulus inhibited this inhibition, producing an increment in response amplitude, i.e., disinhibition. The present findings support this interpretation, since the omission of the CS had the same behavioral consequence as the inhibition of the inhibition controlled by the CS, i.e., in both cases the inhibition was removed or reduced. The observed positive relationship between these effects and the number of reinforcements supports the contention that they result from an associative process as, of course, do the findings of Kimble and Ost (1961) that the conventional interstimulus interval function is involved.

## SUMMARY

This study tested the hypotheses that the amplitude of the unconditioned GSR is gradually attenuated by the CS during conditioning and that this reduction in UCR amplitude vanishes when the CS is omitted and the UCS presented alone. Three groups of 10 $S$s received either 4, 8, or 16 reinforced presentations of a tone CS and electric shock UCS in a classical delayed paradigm. Following reinforcement the CS was omitted and 2 UCS alone trials were administered. After the tests on the shock alone, 4 extinction trials of the CS alone were given.

It was found that the amplitude of the unconditioned GSR diminished during reinforcement and that the UCR recovered to approximately its original amplitude when the CS was omitted. The amount of increment produced by omitting the CS was a linear function of the logarithm of the number of reinforcements. These results were interpreted as supporting the notion that a conditioned inhibitory process develops during reinforcement, under the control of the CS, which attenuates the amplitude of the UCR in the presence of the CS. The relationship between these findings and similar findings in a study of the relationship between amount of disinhibition and number of reinforcements was thought to add further support to the interpretation offered.

## REFERENCES

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KIMBLE, G. A., & OST, J. W. P. A conditioned inhibitory process in eyelid conditioning. *J. exp. Psychol.*, 1961, 61, 150–156.

KIMMEL, H. D., & FOWLER, R. L. The relationship between disinhibition resulting from compound stimulus presentation and the number of reinforcements. *Off. Naval Res. tech. Rep.*, 1961, No. 3. (Contract NONR 580—09)

# THE EFFECTS OF DRIVE AND DISCRIMINATION
# TRAINING ON STIMULUS GENERALIZATION [1]

## DAVID R. THOMAS [2]

*Duke University*

In 1959 Hanson reported in this journal a study on the influence of discrimination training on stimulus generalization. Hanson trained pigeons to peck at a key illuminated by a monochromatic light of 550 m$\mu$, under a VI reinforcement schedule. After successive discrimination training, Ss were subjected to a generalization test in extinction, following a procedure introduced by Guttman and Kalish (1956).

Hanson found that time to the discrimination criterion varied as a negatively accelerated decreasing function (S+, S−) difference. Postdiscrimination generalization gradients (PDGs) were higher and steeper than a control gradient and in addition, showed a shift in the peak of responding from the S+ in the direction away from the S−. The extent of this shift varied also as a negatively accelerated decreasing function of (S+, S−) difference.

The present study is an extension of Hanson's work. The Ss were tested on three different discrimination problems at three different levels of food deprivation. The literature is inconsistent with regard to the

influence of drive level on discrimination learning, and no study appears to have been reported using the operant free-responding technique. In this study, Ss were tested for generalization before discrimination training was initiated, and in addition, short generalization tests were interspersed during the course of discrimination training so as to trace the development of the changes in the PDG.

## METHOD

*Subjects.*—The Ss were 54 experimentally naive white carneau pigeons obtained from the Palmetto Pigeon Plant in Sumpter, South Carolina.

*Apparatus.*—A bank of four identical Skinner type key pecking apparatuses was used. Each box had the following internal dimensions: width, 15½ in.; depth, 14¼ in.; height, 14½ in. Walls and ceiling were painted flat black; floors were of unpainted Masonite. The S's key of translucent plastic was exposed through a ⅞-in. circular aperture placed 6½ in. above the floor on one side of the box. Directly below the key in the floor of each box was a 1-in. circular aperture through which Ss had access to the food magazine. The magazine, which was operated by a motor-driven cam, allowed S approximately 4 sec. of access to food during each cycle. Between cycles, the food was lowered beyond S's reach. A Plexiglas light fixture with a 15-w. bulb, placed directly above and in front of the floor opening illuminated the opening and the magazine for the duration of each cycle.

Aside from the stimulus light and the magazine light during magazine cycles, the boxes were in darkness throughout the experiment. One box, Number 1, was set aside for generalization testing. The source of illumination for the key in this box was a Bausch and Lomb diffraction grating monochromator, Model 33-86-40, equipped with a 108-w., 6-v., ac ribbon filament tungsten lamp. Boxes 2, 3, and 4 were used for training

purposes only. The keys of these were illuminated by 100-w., 120-v., ac projection lamp. Positive and negative stimuli were provided by Bausch and Lomb interference filters. Each box was equipped with a "positive filter," (transmission peak 550 m$\mu$) and a "negative filter," (590 m$\mu$ for Box 2, 570 m$\mu$ for Box 3, and 558 m$\mu$ for Box 4). The brightness of the filter colors was matched to the monochromator colors by the addition to the filters of an appropriate number of pieces of exposed film (Polaroid No. D 77843). The matching was accomplished with the aid of a photomultiplier tube and its associated amplifier.

The three training boxes were set up to change stimuli automatically according to a prearranged schedule punched on a tape run on a Gerbrands programer. Throughout the entire experiment, masking noises were supplied to all boxes by a Grason-Stadler noise generator, Model 901.

*Procedure.*—Upon arrival at the laboratory, all $S$s were weighed, individually caged, and allowed free access to food and water. Throughout the entire experiment free access to water was always available in the home cages. After 4 to 10 days of free feeding, a stable weight level was achieved by all $S$s and food deprivation was begun. The $S$s were randomly assigned to three body weight levels, 80% of ad lib. weight, 70%, and 60%, and to the three different discrimination problems, thus creating nine groups of $S$s, one for each combination of weight level and discrimination problem. Deprivation ceased for each $S$ when the appropriate weight was reached. At this point training was begun.

All $S$s were adapted, magazine trained, and conditioned to key peck according to a set schedule covering 5 days. During this time, the box was in complete darkness with the exception of the key light and the magazine light during reinforcement cycles.

All $S$s were given VI reinforcement for 5 days following conditioning. Thirty 60-sec. stimulus-on periods, alternated with 10-sec. stimulus-off periods, were given each day. The mean interval between reinforcements, not counting stimulus-off periods, was approximately 60 sec, with a range from 4 sec. to 4 min. During stimulus-off periods, a shutter operated, removing the light from the key and leaving the experimental box dark. Reinforcements were never given during a stimulus-off period. One of the five VI sessions for each $S$ was administered in Box 1, the test box, so as to accustom $S$s

to the test box and to the monochromator-produced stimulus.

On the next day, after a 5-min. warm-up period of VI in the test box, each $S$ was subjected to a generalization test in extinction. Eleven different test stimuli were used— 510, 520, 530, 540, 545, 550, 555, 560, 570, 580, and 590 m$\mu$. The 11 test stimuli were randomized within a series and six different random series were presented to each $S$. This resulted in a schedule of 66 stimulus presentations. Each stimulus presentation was for 30 sec. and was followed by a 10-sec. blackout period, during which the number of responses in the preceding period was recorded and the stimulus was changed.

On the day following the (preliminary) generalization test, discrimination training was begun. All three discrimination groups were trained with the same positive stimulus (S+) 550 m$\mu$, and differed only with respect to the negative stimulus (S−). Each discrimination group consisted of 18 $S$s, 6 at each of the three body weight levels. These groups may be designated by the negative stimulus used: 590 m$\mu$ (trained in Box 2), 570 m$\mu$ (Box 3), and 558 m$\mu$ (Box 4). During discrimination training, responding to the positive stimulus was reinforced according to the same VI schedule previously used. Responding to the negative stimulus was never reinforced. The positive and negative stimuli were presented successively in a prearranged random order. Fifteen 1-min. intervals of S+ and 15 of S− were presented each day. All stimulus changes were made during the 10-sec. blackout periods. The 30 stimulus presentations comprised three blocks of 10, and within each block there were 5 positive and 5 negative stimuli. Discrimination training was continued until a criterion of no responding in five successive periods of S− combined with continued responding to S+ was achieved.

During the course of discrimination training, a three-series generalization test of 33 stimulus presentations was administered at the completion of every even numbered daily session of discrimination training. If the discrimination criterion was met on a day on which an interpolated generalization test was scheduled, the test was omitted. On the day following the achieving of the criterion for discrimination, a final generalization test was administered to each $S$. The final test was carried out in the same manner as the preliminary and interpolated tests, with the exception that 12 test series (132 stimulus presentations) were used.
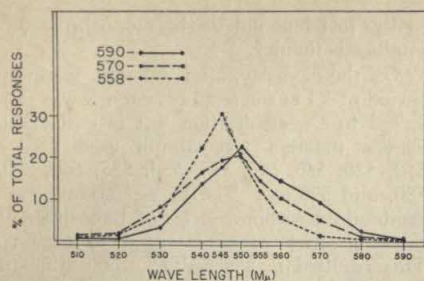
FIG. 1. Postdiscrimination generalization gradients after three different discrimination problems, pooled over three drive levels.

## RESULTS AND DISCUSSION

Data bearing on three different problems will be presented in this report. The first problem concerns the roles of hunger drive level and (S+, S−) difference in determining the rate of formation of a discrimination. The present experiment is fully in agreement with Hanson (1959) in the finding that the rate of formation of a discrimination varies inversely with the physical difference between the stimuli to be discriminated. Analysis of variance indicates that minutes to criterion varies as a function of the (S+, S−) difference ($F = 23.50$, $df = 2/45$, $P < .01$). A beneficial influence of drive level is suggested, but is not demonstrated at a statistically acceptable level of confidence ($F = 2.96$, $df = 2/45$, $.05 < P < .10$). The results seem sufficiently promising, however, to warrant a more direct attack of the problem, without the interpolation (as in the present case) of other experimental treatments (preliminary and interpolated generalization tests). The interaction between the (S+, S−) difference and drive level is not significant ($F < 1$, $df = 4/45$).

In the present experiment deprivation levels down to 60% of ad lib. weight were employed. Thus, the limits of the hunger motive were

assessed. We may therefore conclude that even at its extremes, the hunger drive exercises far less control over the rate of discrimination learning than does the physical difference between the stimuli to be discriminated.

The second problem to be considered here concerns the roles of drive level and the (S+, S−) difference in determining the location of the PDGs. In Fig. 1 are presented the PDGs of the three problem groups pooled over all drive levels. The PDGs show the characteristic displacement of the peak from the S+ an amount inversely related to the (S+, S−) difference. A convenient measure of the location of a gradient is derived by treating it as a grouped frequency distribution and computing the mean value. The difference between the means of the 18 individual mean scores in the three problem groups is significant ($F = 10.74$, $df = 2/45$, $P < .01$). The effect of the (S+, S−) difference on the location of the PDG is mirrored at all three levels of drive. However, the effect of drive level on the means of the final generalization gradients fails to achieve a statistically acceptable level of confidence ($F = 2.56$, $df = 2/45$, $.05 < P < .10$). Neither is the interaction between the (S+, S−) difference and drive level significant ($F = 1.07$, $df = 4/45$). Thus, Hanson's findings with regard to the effect of the (S+, S−) difference on the location of the central tendency of the PDG are replicated at three different levels of drive.

In agreement with Hanson, we have found that both the time to criterion and the amount of displacement of the PDG are inversely related to the (S+, S−) difference. This suggests the possibility that amount of training is the vehicle through which the (S+, S−) difference has its effect on the PDG. The analysis of this ques-

tion was the third major problem to which the present study addressed itself. The administration of a short generalization test to all *S*s after 2 days of training (with the exception of those *S*s which had already met the learning criterion) makes possible an evaluation of the effects of the (S+, S−) difference, with amount of training held constant. On the other hand, the analysis of the series of generalization tests given to all *S*s during the course of discrimination training reveals the role of amount of training with the (S+, S−) difference controlled.

Although only 1 *S* from the 590-mμ group failed to achieve the criterion of discrimination learning within 2 days, 14 *S*s from the 570-mμ group and 16 *S*s from the 558-mμ group could be compared in this manner. For each *S* a "displacement score"



FIG. 2. Mean of the gradient as a function of ordinal position of test.

was obtained by subtracting the mean of the second generalization test from the mean of the first. The means of these displacement scores for the two groups were virtually identical, 5.6 mμ for the 570-mμ group vs. 5.7 mμ for the 558-mμ group. The mean displacement is the same in spite of the fact that in one case the (S+, S−) difference is more than twice that of the other!

In Fig. 2 are presented the means of the preliminary generalization gradients obtained from all *S*s, the means of the second obtained gradient, the third, etc. The data from all groups have been pooled since it appears that neither the (S+, S−) difference nor drive level have any marked effect on the location of the PDG if the amount of training is held constant. Under each of the values is recorded the number of *S*s whose scores are represented in that mean value. It should be noted that the points plotted were not obtained from different groups of *S*s. On the contrary, the 5 *S*s which received five generalization tests were included among the 22 *S*s which received (at least) four generalization tests, which,

TABLE 1

DISPLACEMENT OF THE MEAN OF THE GENERALIZATION GRADIENT WITH TWO SESSIONS OF DISCRIMINATION TRAINING AT TWO DIFFERENT (S+, S−) VALUES

| 570-mμ Group | | 558-mμ Group | |
|---|---|---|---|
| *S* No. | Displacement of Mean (mμ) | *S* No. | Displacement of Mean (mμ) |
| 2 | 12.37 | 15 | 7.91 |
| 5 | 2.39 | 54 | −3.11 |
| 14 | 7.77 | 66 | 4.55 |
| 56 | 14.72 | 51 | 7.90 |
| 70 | −0.62 | 12 | 6.49 |
| 46 | 6.59 | 18 | 8.42 |
| 47 | 7.04 | 44 | 9.52 |
| 59 | 4.77 | 48 | 2.40 |
| 72 | −0.97 | 60 | 4.25 |
| 43 | 2.22 | 67 | 7.91 |
| 40 | 3.21 | 32 | −2.25 |
| 37 | 0.12 | 38 | 2.75 |
| 31 | 7.82 | 25 | 9.64 |
| 62 | 10.27 | 22 | 14.44 |
| | | 29 | 7.07 |
| | | 68 | 3.37 |
| Mean | 5.6 | Mean | 5.7 |

in turn were among the 32 *S*s which received (at least) three generalization tests, etc. The value of *N* continually decreases indicating that all *S*s had at least two generalization tests whereas progressively fewer *S*s had three tests, four tests, etc. The figure shows a negatively accelerated decreasing function. Even as few as two sessions of discrimination training produces a marked shift in the location of the gradient.

The present procedure has made possible the separation of the effects of the (S+, S−) difference and amount of training on the location of the PDG. When this is done it appears that the effects of the (S+, S−) difference are mediated through the amount of training. There is no obvious explanation of the failure of the (S+, S−) difference to have a *direct* effect on the location of the PDG. It is clear, however, that this finding raises serious doubt about the adequacy of a Spence-type theory of inhibition (Spence, 1936, 1937) to account for the process of successive discrimination learning in the operant free-responding situation. It is to be hoped that future studies will suggest an alternative theoretical framework which will have greater predictive and explanatory value within this context.

## Summary

The effects of drive and discrimination training on stimulus generalization were studied in the pigeon. Three groups of 18 *S*s each, maintained at 60%, 70%, and 80% of ad lib. weight, respectively, were trained to respond to a key illuminated by a light of 550 m$\mu$. After 5 days of VI training, *S*s were tested for generalization to stimuli ranging from 500 m$\mu$ to 600 m$\mu$. Then the three groups were subdivided into three discrimination problem groups, 550 m$\mu$ positive for each, but with 590 m$\mu$, 570 m$\mu$, and 558 m$\mu$ negative, respectively. During discrimination training, responding to S+ was VI reinforced, responding to S− was never reinforced. Periodically during the course of discrimination training short generalization tests were given to all *S*s. After the criterion of discrimination was met, all *S*s were subjected to a final generalization test.

The major conclusions were: (*a*) minutes to criterion varies inversely with the (S+, S−) difference; a beneficial effect of drive is suggested but not conclusively demonstrated (.05 < *P* < .10). (*b*) Discrimination training produces a general steepening of the PDG, a lowering of the gradient in the region of S−, and a shift of the central tendency from the region of S+ in the direction away from S−, the amount of shift varying inversely with the (S+, S−) difference. This finding is replicated for all three levels of drive. (*c*) Generalization gradients obtained during the course of discrimination training reveal that the mean of the gradient shifts in a negatively accelerated manner as a function of the amount of discrimination training. The amount of training appears to be the vehicle through which the effect of the (S+, S−) difference on the location of the PDG is mediated. Displacement varies with amount of discrimination training, independent of any direct effect of the (S+, S−) difference.

## References

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hanson, H. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

# THE LEARNING OF RESPONSES TO MULTIPLE WEIGHTED CUES [1]

## STANLEY A. SUMMERS [2]

*University of California, Los Angeles*

In many situations a response is dependent simultaneously on several different cues, i.e., on several potentially useful stimulus attributes. The purpose of this experiment was to analyze the learning of responses to simultaneously presented cues of different validities, in order to determine how much the responses came to depend on each cue. The implications are relevant especially to situations in which individuals react to several influences in reaching a judgment, decision, or interpretation.

Questions about the effects of complex stimuli have been answered by several related types of experimental analysis, such as investigations of stimulus-compound patterning, of probability learning, and of the utilization of multiple cues. In studies of cue utilization, correlations may be taken between each of several cues and the response variable. This correlational approach, which is particularly associated with the work of Brunswik (1956), was used in the present experiment. Brunswik developed the concepts of ecological validity and functional validity. Ecological validity is the correlation between the cue and an environmental variable predicted by it. Functional validity is the correlation between the cue and a response, such as a judgment, directed

toward the environmental variable. Stimulus attributes, then, can be dealt with as joint cues to some criterion variable. A correlational approach to cue utilization provides a method of abstracting from a series of responses the simultaneous effects of several cues.

Brunswik assumed that the functional validity of a cue will correspond, in most cases, to its ecological validity; the extent to which each cue is used should come, through learning, to conform to the extent to which the cue predicts. With correlational analysis Brunswik investigated, for the most part, established patterns of cue utilization, rather than the learning process by which individuals acquire such patterns. He attempted to ascertain the use of cues in settings representative of the natural environment. To isolate the effect of cue validity, however, it seems desirable to establish ecological validities for originally meaningless cues in a laboratory setting and to observe the responses learned to these cues.

An attempt to study the learning of functional validities in a controlled situation was made by R. Goodnow (Bruner, Goodnow, & Austin, 1956). The validity of a cue for predicting a correct category was defined as the relative frequency of the cue's association with that category. The results did not indicate any close correspondence between cue use and objective validity. However, there are limitations in the treatment of this question in terms of frequencies rather than correlations. One limitation is that probability matching, which was considered the criterion of appropriate functional validity, would not have maximized success in the situation and, therefore, did not have the adaptive character of a match of functional to ecological validity in a correla-

COLOR=3 (YELLOW-ORANGE)
ANGLE=6
AREA=7

COLOR=5 (YELLOW-GREEN)
ANGLE=2
AREA=2

COLOR=7 (BLUE)
ANGLE=8
AREA=8
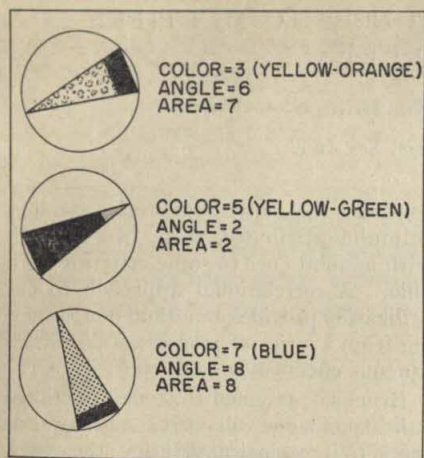
Fig. 1.  Three stimulus forms.

tional framework. Smedslund (1955), dealing less directly with the relation between ecological and functional validity, approached the topic of cue learning through a correlational framework. He established joint visual cues to a predicted variable. A correlation was imposed between each of the cues and the correct response, giving each cue a different validity. The main conclusion was that $S$s do learn "to utilize many probabilistic cues simultaneously . . ." (Smedslund, 1955, p. 26). In Smedslund's design, however, a particular cue was always associated with the same validity, so that the effect of validity could not be studied independent of the effects of saliency or other cue characteristics.

In the present study, correlations were imposed between each of three simultaneously presented visual cues and a variable, correct line length, whose magnitude varied with the magnitude of all three cues. Each of the experimentally imposed correlations will be called a *cue weighting*. Correlations were taken between correct line length and $S$'s response throughout a learning session, and between each of the individual cues and $S$'s response. These empirically

derived correlations will be called *response weightings*.

It was expected that the order of utilization of the three cues would come to conform to the order of cue validity. Specifically, it was predicted that the response weightings would be different and would be ranked in the order of the cue weightings, and that the magnitude of the response weightings would approach that of the cue weightings.

## METHOD

*Subjects.*—The $S$s were 30 members of a ninth grade class, 20 boys and 10 girls. This was a special class composed of students of high academic performance.

*Materials.*—The stimulus materials were 384 geometrical forms photographed on 35-mm. Kodachrome slides and projected on a screen from an automatic projector. Each slide bore a number giving its position in the standard order of presentation. The projected form was approximately 2 ft. in diameter. Each form consisted of an isosceles triangle on a circular white background. A portion of the triangle was black; the remainder was of some hue. Three characteristics of these stimuli varied from slide to slide. These were color, angle, and area. That is, the hue of the triangle, its orientation with respect to the vertical, and the proportion of the triangle that was colored rather than black. Each of the three characteristics varied over eight values; these were designated 1 through 8 (see Fig. 1).

A hue of red was assigned Value 1. The other hues and their values were red-orange (2), yellow-orange (3), yellow (4), yellow-green (5), green (6), blue (7), and violet (8). The two equal sides of the triangle were the long sides; the ratio of long to short was 5:2. If the point opposite the short side is regarded as an indicator, the orientation of the triangle can be described in terms of degrees of clockwise rotation from the 0° (12 o'clock) position. Values were assigned to these positions as follows: 22.5° (1), 67.5° (2), 112.5° (3), 157.5° (4), 202.5° (5), 247.5° (6), 292.5° (7), 337.5° (8). The triangles varied along the dimension of area from almost completely black (1), in equal increments of altitude, to almost completely colored (8). The black strip always paralleled the short side of the triangle. The values (1 through 8) of the

three dimensions varied independently of one another. A standard stimulus series was presented to all *S*s.

The response materials consisted of booklets of paper on which were printed horizontal lines approximately 7 in. long. Each line was an elongated arrow, pointing toward the right. The first 2 pages of a booklet contained 64 lines, numbered 1 to 64, corresponding to numbers on the first 64 slides. The next 16 pages contained 256 bracketed pairs of lines, each pair given a number corresponding to the number in the slide sequence. The lower line of each pair was completely covered by a cardboard strip, which was lightly pasted at both ends. Hidden by this strip was a small red mark which designated a correct length for that line. The last 2 pages of the booklet were identical to the first 2.

*Procedure and design.*—All *S*s were tested in a single session of approximately 2 hr. Two *E*s were present throughout the experiment. The instructions were a fairly straight-forward description of the task and its significance. The following is an abstract of the instructions.

The experiment will investigate how we learn to make certain kinds of judgments. You will see a series of slides projected on a screen—like this. Each of these forms contains the information you need to tell you how to mark off a particular length along an arrow on answer sheets like this. You will just make a small dash, measuring from the left, for how long you think the line should be. Sometimes it will go here, sometimes there. Each slide contains all the information you need to mark off the correct length. Your task in the experiment is to learn how to use this information. At first you will get no help in estimating how long the line should be. Later you will be able to find out how long the line should be after making your own guess. The answer sheets for most of the experiment will have two arrows attached like this. The correct answers are covered by paper strips. You can make a mark right on the line over the strip and then pull down or tear back the strip to see the correct answer. Then you can compare this answer with the mark you made and with the picture. Be careful not to look at the correct mark until after you make your own. Each design on the slide tells you how long the line should be by the features that change from slide to slide. Which way the figure points, which color is on the slide, and the amount of color are all clues to how long the line should be,

and just how, you will have to learn during the experiment. Using all these features at the same time you may learn to get the line length just right. This is difficult but the closer you can come the better. The first two pages are a preliminary measurement before you get help in estimating how long the line should be. The last two pages, also, do not have correct answers marked. You will see each slide only a short time so you will have to judge pretty much by your first quick impression.

A series of 384 slides was presented; each slide was projected for 10 sec., with a 2-sec. interval between slides. During each exposure *S* viewed the slide, made a mark on the upper line of the corresponding line-pair, tore back the cardboard strip, and compared the correct line length marked on the lower line with his mark and with the stimulus pattern. For the first 64 and the last 64 trials, where no correct length was given, *S* simply viewed the stimulus and marked the line. This provided a measure of operant level and an extinction series. A 10-min. rest period was given in the middle of the presentation.

The stimulus series of 384 figures was composed of six blocks of 64 each. The last block (E) was identical to the first (O). Within each block of 64, each of the eight values of each cue appeared equally often in combination with each value of the other two cues. Each block was based on one of five different latin squares, selected from a complete set of 8 × 8 latin squares (Fisher & Yates, 1948, p. 63). The effect of this arrangement was to provide a zero correlation between any two cues within each block and to make almost all stimuli different, aside from the extinction series. Within each block the order of the various combinations of the three cues was randomized, and the distribution of values for each cue was rectangular.

The relation of the cue values to the corresponding series of correct line lengths was defined in the following way. Correct line length was made a function of the three cue values, as expressed by the equation $CLL = 2C_1 + 1.5C_2 + C_3$, where $CLL$ is correct line length, and the $C$s are values of the three cues. The coefficients of the $C$s determine the cue weightings. Since the minimum cue value was 1, and the maximum 8, correct line length varied from 4.5, when all cue values were 1, to 36, when all cues had the value 8. The relation between each cue and correct line length can also be expressed in terms of a correlation. Since the cues were uncorrelated, the correlations of the three cues

TABLE 1

Experimental Conditions: Weightings Assigned Each Cue in Determining Correct Line Length

| Condition | Cue Type and Weighting | | |
|---|---|---|---|
| | .74 | .56 | .37 |
| 1 | Color | Angle | Area |
| 2 | Color | Area | Angle |
| 3 | Angle | Color | Area |
| 4 | Angle | Area | Color |
| 5 | Area | Color | Angle |
| 6 | Area | Angle | Color |

with correct line length were in the ratio 2:1.5:1. Since the square of a correlation corresponds to the proportion of total variance predicted, it is possible to calculate the value of the correlation coefficient for a cue by the equation:

$$(2r)^2 + (1.5r)^2 + r^2 = 1.00$$
$$r = .371$$

Thus the three correlations, or cue weightings, were .371, .557, and .743. The procedures provided $S$ with all the information necessary to make a correct response if he used each cue correctly.

In order to ascertain the effect of various cue weightings independent of other properties of the cue, six experimental conditions were provided. The design was completely counterbalanced with respect to cue type (color, angle, area) and cue weighting and their combinations (see Table 1). These six conditions were six different sets of correct line lengths for the single series of slides. The procedure for the various conditions differed only in the construction of the response booklet, that is, in the correct line lengths. All six experimental groups participated in the experiment together, being presented with the stimulus series simultaneously. The $S$s were randomly assigned to experimental conditions. Data were obtained from 33 $S$s; data from 1 randomly chosen $S$ in each of three conditions were discarded to provide 5 $S$s in each condition.

RESULTS

For each block of 64 presentations, four Pearson correlation coefficients were computed for each $S$. These were the correlations between his response and each cue, as well as between his response and correct

line length. The correlations were averaged over $S$s by converting each correlation to Fisher's $z$, averaging the $z$'s, and converting the average to a Pearson $r$. Curves showing average response weightings for each cue weighting are presented in Fig. 2. The vertical position of the curves corresponds to the order of the cue weightings.

An analysis of variance was performed on $z$ transformations of the response weightings for the learning blocks, the four blocks in which correct line length was presented to $S$. The main question to be answered by the analysis is whether $S$s were learning to respond differentially to the different cue weightings; that is, whether there was a significant difference in vertical placement of the curves. The analysis (Lindquist, 1953, Type VI) is summarized in Table 2. The $F$ for cue weighting is significant beyond the .001 level. This result confirms the prediction that the response weightings would differ with the cue weightings.

In addition, there is an over-all trend toward increasing cue utilization over these four blocks. The block effect is significant beyond the
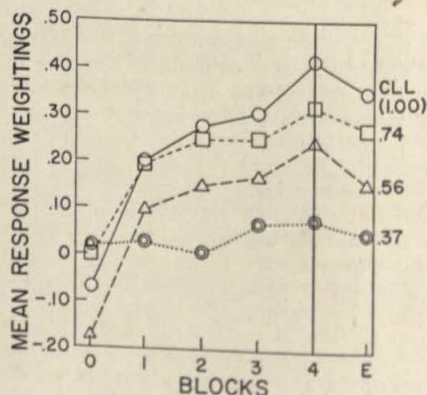


Fig. 2. Mean response weightings (correlations of cue with response) for different cue weightings. (Correct line length was presented in Blocks 1 through 4.)

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSE
WEIGHTINGS USING A $z$
TRANSFORMATION

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | | | |
|   Conditions (C) | 5 | .1382 | <1 |
|   Error (b) | 24 | .2529 | |
| Within Ss | | | |
|   Cue weighting (W) | 2 | 1.3726 | 8.96** |
|   Blocks (B) | 3 | .1858 | 5.40* |
|   W × B | 6 | .0233 | 1.03 |
|   W × C | 10 | .2071 | 1.35 |
|   B × C | 15 | .0154 | <1 |
|   W × B × C | 30 | .0276 | 1.22 |
|   Error (w) | | | |
|     Error₁ | 72 | .0344 | |
|     Error₂ | 48 | .1531 | |
|     Error₃ | 144 | .0227 | |

\* $P < .005$.
\*\* $P < .001$.

.005 level. In order to ascertain whether there is a trend toward increasing utilization of each separate cue, the block effect was tested for each cue weighting separately. The block effect was significant for the .74 weighting ($P < .05$, $F = 2.79$, $df = 3/72$) and for the .56 weighting ($P < .025$, $F = 3.34$, $df = 3/72$), but not for the .37 weighting ($F < 1$). The form of the learning curves does not differ for different cue weightings; the W × B interaction is not significant.

The question of the final magnitudes of the response weightings could not be answered since there was no indication that learning had reached a limit. It is possible, however, to take the observed correlation of response with correct line length on a given block and to compute what the three response weightings would be if they contributed to this total correlation in the expected ratio 2:1.5:1. The response weightings computed on this basis are presented in Table 3, along with the corresponding observed weightings. Table 3 indicates that cue utilization,

throughout the learning trials, is roughly proportional to cue validity, although the highest-weighted cue is being used slightly more, and the lowest slightly less, than would be predicted on the basis of proportionality.

DISCUSSION

The results indicate that Ss responded simultaneously and differentially to the multiple cues. The differential response is shown by the highly significant cue weighting effect. Simultaneous response to the multiple cues, rather than use by each S of only a single cue, is indicated by the fact that this weighting effect transcends the S × Weighting interaction contained in the error term (Error₂). Also, once Ss are exposed to correct line length, the correlation of response with correct line length is higher than with the most heavily weighted cue (see Fig. 2), indicating successful use of more than the most valid cue by itself. Individual Ss must have responded to more than one cue during any given block of trials. The curves suggest that there is some response to the lowest-

TABLE 3

OBSERVED MEAN RESPONSE WEIGHTINGS (O)
COMPARED WITH RESPONSE WEIGHTINGS
COMPUTED ON THE BASIS OF
PROPORTIONALITY (P)

| Cue Weighting | Response Weighting | | | |
|---|---|---|---|---|
| | Block | | | |
| | 1 | 2 | 3 | 4 |
| CLL | .203 | .271 | .313 | .418 |
| .74 | | | | |
|   P | .15 | .20 | .23 | .31 |
|   O | .19 | .24 | .25 | .32 |
| .56 | | | | |
|   P | .11 | .15 | .17 | .23 |
|   O | .10 | .15 | .17 | .24 |
| .37 | | | | |
|   P | .08 | .10 | .12 | .16 |
|   O | .04 | .01 | .06 | .07 |

weighted cue, although in the statistical test this was not rising significantly. In general the *S*s made an appropriate use of the cues available.

The importance to adaptive behavior of appropriate response to multiple cues has been emphasized for various situations: in terms of the learning of functional relations (Miller, 1959), for prediction of judgmental behavior (Johnson, 1955), with application to clinical appraisal (Hammond, 1955; Hoffman, 1960). Important differences of emphasis are implied in the choice of any one methodological framework for the study of cue utilization. In particular one may contrast the orientation implied by a correlational framework with that implied by an event frequency framework, such as is used in studies of probability matching. Both frameworks may be used to study adaptation to environmental uncertainties. The event frequency situation provides for the success of a given response to be probabilistic. Usually the best adaptation possible is consistent choice of the most frequent alternative, with probability matching implying less than optimal adjustment. In a correlational framework, the probabilistic character of the individual cues does not necessarily adhere to the cues in combination, which may be unequivocally related to the criterion as they were in the present study. In terms of behavior, the matching of response weighting to cue weighting would imply fully optimal adjustment to the situation; there may be consistently accurate responses on the basis of individually limited cues. A correlational framework, then, emphasizes cue utilization as a process which tends to remove uncertainty. Much of our judgmental and discriminative behavior is characterized by a sufficient degree of sureness and adaptiveness to be best studied within such a framework.

## SUMMARY

This experiment investigated the relation between the objective validity of certain cues and the extent to which these cues were used.

The purpose was to analyze the learning of responses to multiple cues of different validities, in order to determine how much the responses came to depend on each cue. The independent variable was the correlation imposed between a criterion and each of three simultaneously presented visual cues. The dependent variable was the correlation between the cues and *S*s' responses.

The *S*s were 30 members of a ninth grade class. The stimulus materials were geometric forms. Three characteristics of these stimuli varied, acting as cues for the prediction of a line length. These characteristics were based on the color, angle, and area of parts of the forms, and were counterbalanced to establish independent cue validities. The cue validities of .74, .56, and .37 together permitted a perfect prediction of line length. During four blocks of 64 trials each, *S*s responded and then were presented with the correct line length.

Successful cue utilization increased during the learning session. The *S*s responded to different cues simultaneously, and the extent to which cues were used differed with validity. Cue utilization was roughly proportional to cue validity throughout the learning trials.

## REFERENCES

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver & Boyd, 1948.

Hammond, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.,* 1955, 62, 255–262.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.,* 1960, 57, 116–131.

Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

Lindquist, E. F. *Design and analysis of experiments.* New York: Houghton Mifflin, 1953.

Miller, N. E. Liberalization of basic S-R concepts. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

Smedslund, J. *Multiple-probability learning.* Oslo: Akademisk Forlag, 1955.

# STIMULUS-RESPONSE CONTIGUITY IN CLASSICAL AVERSIVE CONDITIONING [1]

R. A. CHAMPION

*University of Sydney*

Stimulus-response contiguity is widely accepted as a necessary, if not sufficient, condition of learning. In particular, it has been suggested that even though reinforcement may be required for complete learning, a response which is being learned cannot be reinforced until it has begun to occur and that S-R continguity may operate as a crucial factor in *initiating* any learning process (Mason, 1959). In classical conditioning, degree of S-R contiguity is represented in the temporal interval between the CS and the UCR, and the following experiment was designed to test the hypothesis that, early in conditioning, performance is proportional to degree of CS-UCR contiguity. In order to make this test a short-latency response (eyeblink) and a long-latency response (GSR) were conditioned simultaneously in each S and the time relations between CS and UCR were varied so as to permit greater contiguity of eyeblink and CS in one group of Ss and greater contiguity of GSR and CS in a second group. This type of arrangement is represented in Fig. 1; with an air puff as the UCS, the CS (tone, T) may be presented near the unconditioned eyeblink as in the case of T1, or near the unconditioned GSR as with T2. In terms of the contiguity hypothesis it was predicted that the eyeblink would condition better than the GSR with T1 whereas the reverse would hold with T2, i.e., that there would be inter-

action between the CS-UCS interval and the latency of the UCR as independent variables affecting conditioning. By the same token it was expected that the early conditioning of the eyeblink would be better with T1 than with T2 but that T2 would produce the superior performance in the case of the GSR.

The experiment provided an opportunity to confirm an earlier finding that backward conditioning produced some true learning of the long-latency GSR (Champion & Jones, 1961). It is apparent from Fig. 1 that UCS-CS trials must be used in the conditioning of the GSR if optimum CS-UCR contiguity is to be achieved and this arrangement was therefore employed in the second of the two groups described above. After these two groups had been trained, however, the question arose as to the real meaning of "backward conditioning." The accepted definition seems to be put simply in terms of the presentation of the UCS before the CS, but this does not guarantee that the UCR will precede the CS. In fact, in the second
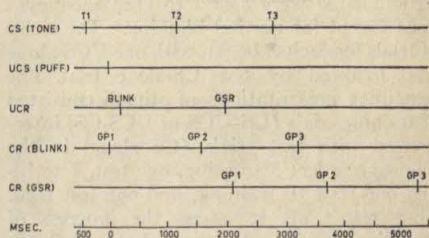


FIG. 1. Schematic representation of time relations between stimuli and responses in the three groups of the experiment. (T1, T2, and T3 represent the location of the CS in Groups 1, 2, and 3, respectively.)

arrangement described above, favoring contiguity of GSR and tone (T2 in Fig. 1), the CS still precedes the UCR. A third group was therefore trained with an even longer backward UCS-CS interval so as to ensure that the unconditioned GSR preceded the CS (T3 in Fig. 1). An attempt was made to set this interval so that degree of contiguity was nevertheless approximately the same as in the second group, i.e., so that CS-UCR contiguity matched UCR-CS contiguity. The use of this third group allowed a test of the possibility that R-S and S-R contiguity are equally effective.

## METHOD

*Subjects.*—The *S*s were 49 volunteers from courses in psychology at the University of Sydney.

*Apparatus.*—The general experimental situation and the apparatus for measuring the GSR have been described elsewhere (Champion & Jones, 1961). The eyeblink was measured with the system developed at the State University of Iowa laboratory, i.e., a microtorque potentiometer linked mechanically to *S*'s right eyelid and electrically to an ink-writing recorder through a dc amplifier. The CS was a 2000-cps, 90-db. tone of 50-msec. duration delivered to *S* through headphones. The UCS was a 3.0-psi air puff applied to *S*'s right eye through a .062-in. diameter tube and set at 50 msec. duration by means of an ac solenoid valve. All time intervals except the intertrial interval were controlled with electronic timers.

*Procedure.*—In the course of general instructions *S*s were asked to keep their eyes open except for normal blinking. The series of trials for each *S* began with one UCS-alone trial followed by one CS-alone trial. The remaining presentations of stimuli consisted of training trials (CS-UCS or UCS-CS) interspersed with test trials (CS alone) in the following order: 3 training, one test, 7 training, one test, 10 training, and one test trial. Thus test trials to assess the progress of conditioning were administered after 0, 3, 10, and 20 training trials. The training series was limited to 20 trials because the hypothesis under test dealt only with the early stages of conditioning. The intertrial interval averaged 30 sec. and varied between 25, 30, and 35 sec. in prearranged random order. No ready signal was given at any stage.

As represented in Fig. 1, Groups 1 and 2 were conditioned with a CS-UCS interval of 400 msec. and a UCS-CS interval of 1200 msec., respectively. Group 3 was then trained with a UCS-CS interval of 2800 msec. A conditioned eyeblink was defined as occurring on test trials with a pen deflection of 1 mm. or more within the interval 200–500 msec. following the onset of the CS. No attempt was made to exclude "voluntary responders." A GSR on test trials was taken to be any response occurring in the interval 1–4 sec. after the CS. There was some adaptation of the GSR to the air puff and the results of 4 *S*s who failed to give a response on 6 or more of the 20 training trials were discarded. This left 15 *S*s in each of the three groups. There were 7, 8, and 6 women in Groups 1, 2, and 3, respectively; a survey of the results showed that the men gave insignificantly more CRs of each type compared with the women.

## RESULTS

In view of the predicted interaction between the length of the CS-UCS interval and the latency of the UCR it was necessary to make a direct comparison between the performance of the eyeblink and the GSR. For this purpose a technique was used which had already proved effective (Mason, 1959) and which allowed a measure of the frequency of the conditioned GSR. For each *S* the GSR scores (change in conductance in micromhos) obtained on test trials during training were expressed as a ratio of the response on the first test trial, before training; this transformation overcame individual differences in general sensitivity of response. In the few cases where there was no response on the first test trial the least measurable response of 0.1 micromhos was assumed. The transformed scores, ranging from 0 to 35.0, were pooled and found to have a median of 2.6. Each value at or above 2.6 in the transformed scores was then taken as a conditioned GSR.

The eyeblink and the GSR were conditioned simultaneously in each $S$ and it was possible to test for the predicted interaction by comparing across groups the proportion of $S$s giving more conditioned eyeblinks than conditioned GSRs on the test trials during training. These data are set out in Table 1; the result for Groups 1 and 2 was in accord with the prediction and a Fisher exact probability test (Siegel, 1956), applied with cases of equal frequency of response omitted, showed the interaction to be significant ($P < .01$ for a two-tailed test). Comparison of these results with the data from Group



FIG. 2. Performance curves for eyeblink conditioning.

TABLE 1

NUMBER OF $S$s GIVING MORE, EQUAL, AND FEWER CONDITIONED EYEBLINKS COMPARED WITH CONDITIONED GSRs ON TEST TRIALS DURING TRAINING

| Outcome (CRs) | Group 1 ($N = 15$) | Group 2 ($N = 15$) | Group 3 ($N = 15$) |
|---|---|---|---|
| Eyeblinks > GSRs | 9 | 0 | 3 |
| Eyeblinks = GSRs | 3 | 4 | 2 |
| Eyeblinks < GSRs | 3 | 11 | 10 |

3 showed the interaction to be significant for Groups 1 and 3 ($P < .05$) but not for Groups 2 and 3.

The other aspect of the predicted interaction involved within-response comparisons over the various CS-UCS intervals. Performance curves for the eyeblink CR are shown in Fig. 2. The performance of Group 1 was consistently superior to that of Group 2, as predicted, but the application of Fisher tests showed that a statistically significant difference did not emerge until 20 training trials ($P < .01$); at 10 trials the difference approached significance ($.10 > P > .05$). The same relationship held between Groups 1 and 3 but there was no significant difference at any stage

between Groups 2 and 3. An overall test at 20 trials proved significant ($\chi^2 = 12.09, P < .01, df = 2$). These results conform to the general finding that a 400-msec. interval is almost optimum for the eyeblink and that backward conditioning of this response is ineffective.

The more important within-response comparison concerned the GSR, and performance curves for the conditioning of this response are shown in Fig. 3. As predicted, Group 2 was consistently superior to Group 1, but according to Fisher tests the difference was only statistically significant at 3 trials ($P < .05$). Groups 1 and 3 did not differ significantly at any stage. An
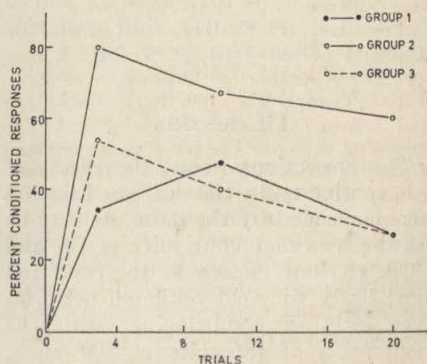


FIG. 3. Performance curves for GSR conditioning.

overall test at 3 trials gave a significant result ($\chi^2 = 6.66$, $P < .05$, $df = 2$). More sensitive measures than frequency of response were available for the GSR and the transformed amplitude scores were studied for further information. Nonparametric statistical tests were used because of the marked positive skew in the distribution of these scores. The results given above were confirmed; in addition, the application of a $U$ test at 20 trials showed that the difference between Groups 1 and 2 approached significance ($U = 68$, $.10 > P > .05$). At this stage there was a significant difference in amplitude of response between Groups 2 and 3, with the former superior ($U = 51$, $P < .05$), but not between Groups 1 and 3. An overall test applied to the three groups at 20 trials proved significant ($H = 6.99$, $P < .05$, $df = 2$).

As a check on the possible introduction of some artifact in the transformation of the GSR scores, Mood's test of trend (McNemar, 1955) was applied to the raw change-in-conductance scores for each group. The observed trends in these data were virtually identical with those in the frequency data, depicted in Fig. 3; they were significant for Groups 2 ($\chi^2 = 11.8$, $P < .01$, $df = 3$) and 3 ($\chi^2 = 11.4$, $P < .01$), but not for Group 1 ($\chi^2 = 2.20$, $P > .50$).

### DISCUSSION

The predictions about the effects of S-R contiguity in classical conditioning were confirmed by the data. It may be asked why exact contiguity of CS and UCR was not chosen as the condition most likely to yield optimum performance early in conditioning; with the assumption of latencies of 200 msec. and 2000 msec. for the unconditioned eyeblink and GSR, respectively, this would have been achieved with the use of corresponding backward UCS-CS intervals. A pilot study conducted along these lines before the present experiment produced results exactly in keeping with the contiguity hypothesis, but the level of conditioning in both groups was so low that the interaction was not statistically significant. Poor conditioning with a short backward interval is an established result for a short-latency response (e.g., Wolfle, 1932) and the results of Groups 2 and 3 in the present experiment suggest that there is a decrease in the effectiveness of GSR conditioning in early stages with a backward interval somewhere between 1200 msec. and 2800 msec., possibly of the order of 2000 msec. An obvious inference to be drawn from these considerations is that the required contiguity lies not between CS and UCR as overt events but rather between corresponding physiological processes, lagging behind the CS in stimulus reception and preceding the UCR in response evocation.

It is tempting to suppose that the CS must precede the UCR for contiguity to be effective. This would explain why the conditioning of the eyeblink did not occur in Groups 2 and 3, and why Group 3 was inferior to Group 2 and comparable with Group 1 on some counts in the conditioning of the GSR. The supposition is contradicted, however, by the comparable performance of Groups 2 and 3 with respect to the relative frequency of the two types of CR (Table 1) and by the presence of a significant trend in the GSR conditioning of Group 3. Both these latter results are probably due to the early rise in the performance of Group 3, at three training trials. Contiguity appears to have had some effect when the UCR preceded the CS (Group 3), but the performance was inferior to and not as sustained as that obtained with a comparable degree of CS-UCR contiguity (Group 2). Such an outcome could be due to two factors which are not mutually exclusive. First, if allowance is made for the physiological processes

delineated above, then the contiguity between stimulus reception and response initiation would be greater in Group 2 than in Group 3 (Fig. 1); under the conditions of Group 3, of course, some physiological mechanism must be found which permits contiguity to act when response initiation precedes stimulus reception. Second, if delay of reinforcement is represented in the interval between CR and UCS, then it is greater in Group 3 than in Group 2 and performance in the former group should not have been so sustained as in the latter group for that reason; this consideration brings with it the problem of "backward" reinforcement when the UCS precedes the CR. More generally, it may seem improper to invoke the effects of reinforcement if attention is limited to the initial stages of learning, where contiguity has been taken as the prime factor operating. Reinforcement may take effect, however, as soon as the learned response appears, no matter how weakly, and the precise separation of this effect from that of contiguity awaits the use of some more refined technique than mere reference to early and late stages of learning.

Attention should be drawn to two other features of the data. There is some conflict with the results of the previous study (Champion & Jones, 1961), where the backward conditioning of the GSR (with a UCS-CS interval of 500 msec.) was inferior from the outset to forward conditioning (with a CS-UCS interval of 500 msec.) whereas the performance of Group 2 was superior to that of Group 1 in the present study. The discrepancy might be accounted for in terms of differential degrees of contiguity and reinforcement, but the differences in procedure in the two experiments are too great to permit this explanation to be pursued with any confidence. A second point of interest was the appearance in the present data of a statistically significant contiguity effect early in training with the GSR but relatively later with the eyeblink. This result would emerge if greater contiguity was achieved by chance in Group 2 with the GSR than in Group 1 with the eyeblink. It should be noted that the only set of conditions favoring performance on the grounds of both contiguity and reinforcement was that prevailing in Group 1 for the eyeblink and that only under these conditions did the performance curve rise throughout its course.

## SUMMARY

A test was made of the hypothesis that CS-UCR contiguity is an important factor in the initial stages of classical aversive conditioning. The short-latency eyeblink and long-latency GSR were conditioned simultaneously in each $S$, with tone as CS and air puff as UCS. A forward CS-UCS interval of 400 msec. and a backward UCS-CS interval of 1200 msec. were used in separate groups to favor contiguity of CS with eyeblink and GSR, respectively. The acquisition of the eyeblink was superior with the forward interval and inferior with the backward interval when compared with that of the GSR. Within-response comparisons showed that the eyeblink conditioned better with the forward interval whereas the GSR conditioned better with the backward interval. To test the effect of UCR-CS contiguity in the case of the GSR a third group was trained with a backward UCS-CS interval of 2800 msec.; this condition produced an initial rise in performance which was not sustained. The results of the experiment were interpreted as supporting the hypothesis.

## REFERENCES

CHAMPION, R. A., & JONES, J. E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–62.

McNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

MASON, J. E. The joint action of contiguity and reinforcement in classical conditioning. Unpublished doctoral dissertation, University of Sydney, 1959.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

WOLFLE, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

# ASSOCIATIONS, SETS, AND THE SOLUTION OF WORD PROBLEMS [1]

MIRIAM A. SAFREN [2]

*Johns Hopkins University*

Investigators have used word problems (anagrams and skeleton words) to explore the "category set," which can be defined as a readiness to respond to words belonging to a common category or class, and can be measured by the speed or frequency[3] with which certain responses occur. For example, Starch (1911) and others demonstrated the existence of the category set when *S*s solved problems made from lists of words belonging to a common class more rapidly than problems made from random word lists. Later, Rees and Israel (1935) and others measured the strength of the category set by the frequency with which *S*s solved multiple solution problems as category words after practice on unique solution problems which could be solved only as words of the selected category. Past *E*s reported that telling *S*s the category name did not always have a consistent effect; what they are surer of is that the set arises from verbal context during problem solving. Unfortu-

nately, the past *E*s selected the category words and category name on an a priori basis, and none of them systematically investigated the verbal context triggering the set. The present study aims to show that associative strength between words from which problems are made is a pertinent variable to manipulate in the operation of the category set in anagram solving.

This approach stems from work in free recall showing that associative strength between words presented to *S*s significantly determines *S*s' verbal productions. Additional support for this approach is given by Mayzner and Tresselt (1958) who suggest that processes in anagram solving may be described by laws similar to those in verbal recall.

Especially relevant to the problem investigated here is a study by Deese (1959) who investigated the effect on recall of presenting *S*s with lists of words all of which are associated, with the average strength of association among the group expressed in an index he called interitem associative strength. He found that in "organized" lists of words high on this index, *S*s recalled more and had fewer verbal intrusions. Postulating that free recall is in part free associating, Deese explained that *S*s recalled more with the organized lists because there was a greater probability of *S*s' associations actually being listed items.

The present study extends Deese's model of free recall to anagram solving and hypothesizes that the category set in anagram solving operates by associations between words selected by *E*s as belonging to a given category. On the basis of this hypothesis,

## TABLE 1
### The Six Organized Lists Used in the Experiment

| List 1 | | | List 2 | | | List 3 | | |
|---|---|---|---|---|---|---|---|---|
| Word | Anagram | Labels (Cond. O-L) | Word | Anagram | Labels (Cond. O-L) | Word | Anagram | Labels (Cond. O-L) |
| COMMAND | CANDOMM | | WHISTLE | WHELIST | | CHAIR | RAICH | |
| ORDER | REDOR | Military | TRAIN | TARNI | Railroad | SOFT | FOST | Comfort |
| ARMY | RAMY | Armed Forces | NOISE | NESOI | Station | SOFA | OFAS | Relaxation |
| OBEY | OYEB | Discipline | SOUND | UNDOS | Sounds | CUSHION | CUNSHIO | Furniture |
| SOLDIER | SODLERI | | SHRILL | SHLIRL | Noises | PILLOW | WOLPIL | |
| NAVY | VANY | | LOUD | OLDU | | COUCH | COCHU | |

| List 4 | | | List 5 | | | List 6 | | |
|---|---|---|---|---|---|---|---|---|
| MILK | LIMK | | DOCTOR | DOROCT | | SQUARE | SAQURE | |
| CREAM | RECAM | BEVERAGES | NURSE | URNSE | Hospital | CIRCLE | CIRLEC | Geometry |
| SUGAR | USGRA | Breakfast | HEALTH | EHALHT | Physical Health | ROUND | DUNRO | Geometric |
| COFFEE | EFECOF | Food | SICK | KISC | Medical Care | CUBE | BECU | Figures |
| SWEET | SEWTE | Taste | MEDICINE | MEDCIENI | | BLOCK | CLOBK | Shapes |
| DRINK | DINRK | | CURE | ECRU | | BALL | LALB | |

the following predictions were made: (*a*) Anagrams made from organized lists of associatively related words will be solved more quickly than anagrams made from random word lists, because in the former, associations called up by solved problems should aid *S*s in the solution of subsequent ones. (*b*) There should be a rapid decrease in solution time with trials when *S*s solve anagrams made from associatively related words as associations called up by solved problems *begin* to aid *S*s in the solution of subsequent ones.

## METHOD

The study consisted of two parts: the selection of stimulus materials and the experiment itself.

*The experiment.*—The experiment consisted of three major conditions. In Cond. O and Cond. O-L groups of *S*s were presented with anagrams made from "organized" lists of words (words that elicit each other in free association). In addition, *S*s assigned to Cond. O-L were told labels describing the list of words from which the anagrams were made. These labels were selected on the basis of normative data gathered prior to the experiment. The labels were associates to the words of the list and in turn elicited them. In Cond. R, *S*s were given anagrams made from "random" lists of words (words with essentially

zero probability of eliciting each other in free association). The experimental design was such that the same sample of words in the same anagram form was used for all experimental conditions. Thus, as many factors as possible were constant for all conditions and whatever effects might occur could be attributed to the influence of associative context.

*Stimulus materials.*—The stimulus materials consisted of 36 anagrams. The 36 words from which the anagrams were made came from 6 organized lists of 6 words each. These lists were constructed after administering several word association tests consisting of 200 items including filler items to groups of from 50 to 100 Johns Hopkins University undergraduates. The instructions given to *S*s in this part of the study were standard word association test instructions.

The following governed the selection of the six words of each organized list: (*a*) the six words should have maximum interitem associative strength, i.e., the words should have a maximum probability of eliciting each other in free association, and (*b*) they should be words from which only unique[4] solution anagrams could be made. The six lists of six words each are shown in Table 1.

Labels describing the organized lists (Cond. O-L) were obtained as follows: A

[4] Two of these words do not form unique solution anagrams. These are SOFA which could be structured as OAFS and CIRCLE which could also be CLERIC. In the anagram form in which they were presented, none of these solutions occurred.

TABLE 2

OVER-ALL MEDIAN SOLUTION TIMES (IN SEC.) FOR SIX ORGANIZED LISTS OF
DIFFERING INTERITEM ASSOCIATIVE STRENGTHS WITH AND
WITHOUT LABELS

| List | Interitem Associative Strength | | No. Words Labels Elicit | | Median Time (No Labels) | Median Time (with Labels) |
|---|---|---|---|---|---|---|
| | Mean | Variance | Mean | Variance | | |
| 1 | 8.0 | 245.07 | 1.11 | 1.32 | 13.65 | 6.60 |
| 2 | 7.4 | 146.86 | 1.38 | .96 | 7.65 | 3.85 |
| 3 | 6.1 | 79.40 | 1.22 | .90 | 10.85 | 2.90 |
| 4 | 5.2 | 95.46 | 1.86 | 1.06 | 3.25 | 1.25 |
| 5 | 8.3 | 137.72 | 2.03 | 1.69 | 16.45 | 2.10 |
| 6 | 9.1 | 183.40 | 1.64 | 1.18 | 3.25 | 1.25 |

Note.—Median time scores for each list are based upon data of 6 Ss who solved six anagrams each.

group of 60 Ss, Johns Hopkins University undergraduates, were given a 6-page booklet containing the six organized lists, one per page. The Ss were told to look at each list of words and to think of a label or labels that described them or the way in which they were similar, and to note this in the space provided. The various labels and the frequency with which they were given were tabulated. The labels selected for each list were those which included 50% or more of the responses. The labels are shown in Table 1.

As the labels of each list were associates to the words of that list, further normative data were obtained to find out if the labels would elicit the words of the list. A sample of 40 Ss, Johns Hopkins University undergraduates, were given a 6-page booklet containing the six sets of labels, one per page. The Ss were told to look at the labels on each page and then to write down the words which the labels called to mind, but to spend no longer than 1 min. associating to each set of labels. Table 2 shows the average number of words of each organized list which the labels of that list elicited. The mean number of words called up by these labels in 1 min. was 10 ($SD = 6.32$).

TABLE 3

ANAGRAM SOLUTION TIMES (IN SEC.) FOR
THE THREE CONDITIONS

| Condition | N | Median | Q | Range |
|---|---|---|---|---|
| R | 36 | 12.2 | 6.2 | 2–67.8 |
| O | 36 | 7.4 | 5.6 | .1–49.2 |
| O-L | 36 | 2.8 | 1.9 | .6–22.4 |

Note.—Time scores on which medians, Qs, and ranges are calculated are the median solution times for each of the 36 Ss tested under each condition.

For Cond. R in which Ss were presented with anagrams made from random lists of words, 36 random lists[5] of six words each were generated as follows: One word at a time was taken from each of the 6 organized lists to form a random list of six words. By a prearranged order, every word of each organized list occurred in 6 of the 36 random lists but never in the context of any other word of the same organized list.

In constructing the anagrams, an attempt was made to control for three factors that affect difficulty of solution (Mayzner & Tresselt, 1958, 1959; Sargent, 1940): (a) the degree of difference of the letter arrangement from the original order in the word, (b) the transitional probabilities between the letters in the disarranged form, and (c) frequency of occurrence of the word in printed matter (Thorndike & Lorge, 1944).

*Subjects.*—A total of 108 Ss from the introductory psychology class at the Johns Hopkins University were randomly assigned, 36 per condition. For Cond. O and O-L there were 6 lists of six anagrams each. Different groups of 6 Ss received each of the 6 lists, and within each group the order of occurrence of the six anagrams was counterbalanced over Ss. For Cond. R, there were 36 lists of six anagrams each. Each of the 36 Ss assigned to this condition received a different list.

*Procedure.*—All Ss were presented with six anagrams typed in capitals on 3 × 5 in. cards and told the following: (a) they would have six anagrams to solve; (b) an anagram is a word with the letters disarranged; (c) they were to look at the anagram, try to determine the original word without the use

[5] The 36 random lists are in the doctoral dissertation on file at the library of the Johns Hopkins University.

of paper and pencil, and then to give the solution orally; (*d*) they would have 4 min. to solve each problem and then would be given the solution; (*e*) they were to work as rapidly as possible as they were being timed on each problem. In the case of Cond. O-L, *S*s were told that the anagrams were made from a list of words related by association, and given the labels that described the list. Solution times were recorded with a stopwatch, and all *S*s were tested individually.

## RESULTS

Because of the nature of the distribution of *S*s' anagram solution times, medians rather than means were chosen to represent *S*'s typical performance in the statistical analysis. Table 3 summarizes the data on solution time for the three conditions showing that medians, *Q*s, and ranges decrease in going from Cond. R to O to O-L. The overall statistical significance of the difference between the three experimental conditions, using the Kruskall-Wallis one-way analysis of variance, is $P < .005$. A comparison of Cond. O with Cond. R by the Mann-Whitney $U$ test yielded $P < .02$. A similar comparison between Cond. O and O-L yielded $P < .003$.

With a significant difference between Cond. R and Cond. O, a detailed analysis was made of the organized lists of differing interitem associative strengths. A Kruskall-Wallis one-way analysis of variance by ranks showed a significant difference between organized lists ($P < .05$). However, the correlation (Kendall's tau) between interitem associative strength index and median solution time for the organized lists was $-.28$ and nonsignificant. Table 2 shows this information.

Since the comparison between Cond. O and Cond. O-L was very significant, Mann-Whitney $U$ tests were done comparing median solution times for the 6 *S*s of each organized list with the comparable 6 *S*s of the same organized list with labels. For Lists 3, 5, and 6 there were significant differences between the two conditions ($P = .021$, $.013$, and $.002$, respectively). For Lists 1, 2, and 4 the differences were not significant ($P = .155$, $.242$, and $.155$, respectively). Therefore, it may be concluded that anagram solution is facilitated when *S*s are given anagrams made from "organized" associatively related word lists. Anagram solution is facilitated still further when in addition, *S*s are told labels describing the organized lists when these labels are (*a*) associates to the words and (*b*) in turn call up the words of the lists.

The overall Trial effect (*S*s under all conditions had six problems to solve), as tested by the Friedman two-way analysis of variance by ranks, was significant ($P < .02$). However, there appears to be an interaction between trials and conditions (see Fig. 1), for the Trial effect is nonsignificant for Cond. R ($.50 < P < .70$) and nonsignificant for Cond. O-L ($.10 < P < .20$). The Trial effect is significant for Cond. O ($P < .05$). The significant Trial effect for Cond O verifies the predicted decrease in solution time with trials when *S*s solve anagrams made from organized lists as associations called up by the context begin to aid *S*s in the solution of subsequent problems.

With a significant Trial effect for Cond. O, analysis was made of the Trial effect for each organized list. Only the Trial effect for List 3 was significant ($P < .02$). However, the sample size for each organized list was only 6.

Although the six organized lists vary little in the mean probability with which the words elicit each other in free association, there is a signifi-
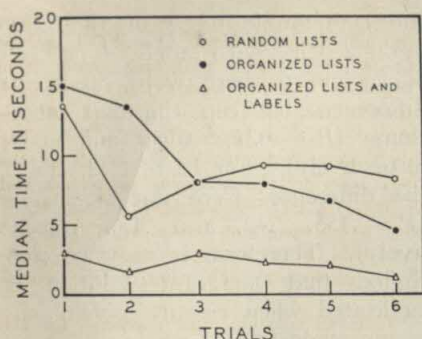
FIG. 1. Median anagram solution time per trial for each of the three experimental conditions. (Each point is based upon the data of 36 *S*s.)

cant difference in the variance about the mean (see Table 2). Bartlett's test for the difference between the $K$ variance estimates was significant at $P < .05$. Furthermore, there seems to be a relationship between the interitem associative strength index and the decrease in solution time with trials for each organized list.[6] Ranking each list by the magnitude of its variance and then ranking each according to the amount of change in solution time with trials results in a correlation (Kendall's tau) of $-.60$, $P = .068$. This suggests that the greater the variance about the interitem associative strength index the less the change in solution time with trials.

A similar correlation (Kendall's tau) between the interitem associative strength of each list and the variance about this value is .60 ($P = .068$). Thus, it appears that organized lists with higher interitem associative strengths had a higher variance about this index which would tend to make the drop in solution time with trials less.

Possibly, not controlling for the variance about the interitem associative strength index was one reason why it was not possible to show a

[6] A table showing the trial effect for each organized list is in the doctoral dissertation.

position relationship between the magnitude of the index and the median solution time for each organized list. It is also possible that other factors inhibiting this relationship were (*a*) the narrow range of interitem associative strengths used, and (*b*) the inability to make anagrams of all organized lists of equal difficulty.

## DISCUSSION

The results support the hypothesis that in the solving of word problems such as anagrams, the category set operates by associations between words selected by $E$ as belonging to a common category. The results also suggest processes in the solving of word problems such as anagrams may be similar to those in verbal recall; for in both verbal recall and anagram solving, associative strength between words significantly determines *S*s' verbal productions.

A suggested model to explain the results is as follows: When *S*s solve a list of anagrams, at least in part, they are calling up words to match the letters presented and sampling from a momentary response pool of available words. After *S*s have determined some of the words from which the problems were made, these words implicitly evoke others which are associatively related. These associations become part of the momentary available pool of words from which *S*s are sampling. Thus, implicitly evoked associations from the solution of preceding problems can facilitate the solution of subsequent problems by converging upon words on the list from which the anagrams were made.

When *S*s are given anagrams made from associatively related words and also given labels for these words, these labels evoke at the *outset* a response pool of available words some of which converge upon solutions to the anagrams. Since both associations elicited by solved anagrams and those called up by labels increase the availability of anagram solution, anagrams are solved most quickly under Cond. O-L. However, since from the outset the labels make

some of the anagram solutions readily available, the Trial effect is not significant for Cond. O-L.

Although couched in different language, the present model of the category set agrees basically with that of Maltzman and Morrisett (1952). The latter explain the set as follows: An anagram belonging to a common class, e.g., "nature," serves as a stimulus for the arousal of other words belonging to the selected class and in addition, a whole hierarchy of responses some of which belong to the selected class, some of which do not. However, since only the selected category responses are consistently reinforced in the experiment, these become dominant through mediated generalization.

Maltzman and Morrisett do not explicitly describe the mediated generalization mechanism. However, in the present model, the Ss' pool of associatively related words implicitly evoked by solved anagrams could be regarded as the mediated generalization mechanism. This pool of words would include the category name in so far as it is associated to the words comprising a common class and elicits these words.

The present description of anagram solving in terms of verbal recall and free association may be considered an extension of a model by Deese (1959), who interprets free recall as involving in part free association. Thus, the present study represents an extension over which associative relationships between words plays an important role.

The present research also implies that in future studies on the category set, stimulus materials should be calibrated by existing associative relationships for more precise predictions on the rate of problem solving. For it is the contention of this study that associative strength between words may be considered a measure of *existing response sets* which when activated, facilitate solutions of word problems such as anagrams.

## SUMMARY

This experiment tested the hypothesis that the category set in anagram solution operates by associative strength between words selected by $E$ as belonging to some common class. In Cond. O and O-L, groups of Ss were given anagrams made from "organized" lists of words that elicited each other in free association. In Cond. O-L, Ss were also told labels describing the lists. The labels were associates to the words on the list from which the anagrams were made and in turn elicited these words. In Cond. R, Ss were given anagrams made from "random" word lists (words with essentially zero probability of eliciting each other in free association).

Time to solution was evidence for the operation of the set. It was predicted that (a) anagrams made from organized lists would be solved more quickly than anagrams made from random lists; for in the former, associations called up by the context would aid in problem solution; (b) there would be a significant Trial effect (a decrease in solution time with trials) for the organized condition as associations called up by context began to aid Ss in the solution of subsequent problems. The results verified these predictions and supported the present interpretation of the category set. A description of anagram solving as involving verbal recall and free association was advanced to explain the results.

## REFERENCES

DESSE, J., Influence of inter-item associative strength upon immediate recall. *Psychol. Rep.*, 1959, **5**, 305–312.

MALTZMAN, I., & MORRISETT, L., JR. Different strengths of set in the solution of anagrams. *J. exp. Psychol.*, 1952, **44**, 242–246.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

REES, H. J., & ISRAEL, H. E. An investigation of the establishment of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

SARGENT, S. S. Thinking processes at various levels of difficulty. *Arch. Psychol., N. Y.*, 1940, **35**(Whole No. 249).

STARCH, D. *Experiments in educational psychology*. New York: Macmillan, 1911.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

# SEQUENTIAL INTERFERENCES DEMONSTRATED BY SERIAL RECONSTRUCTIONS [1]

## E. B. COLEMAN

### *Human Resources Research Office*

The findings of experiments on retroactive and proactive inhibition have been generalized to suggest that forgetting that occurs outside the laboratory is caused by interference similar to that found in the traditional laboratory experiment. Recently Underwood and Postman (1960) noted that "As plausible as this argument may seem, it is apparent that it would be highly desirable to specify possible sources of extraexperimental interference, and to specify these in such a way that experimental tests of their influence on forgetting are possible" (p. 74). They suggested two such interferences: "letter-sequence interferences" and "unit-sequence interferences."

Letter-sequence interferences result when a list presented for learning conflicts with English spelling habits. Suppose $S$ is presented with a consonant syllable JQB. Underwood and Postman assume that previously learned spelling habits will make this sequence hard to learn because Q never follows J; nor does B ever follow Q in English spelling. These previously learned spelling habits must be "extinguished" or "unlearned" or "inhibited" before JQB can be learned. Following Briggs

(1954), they also assume that as time passes, the older sequential habits will spontaneously recover and interfere with JQB during a test for retention (proactive inhibition). The older spelling sequences may be used during the delay period also (common sequences being used more frequently than uncommon ones). This will further strengthen them and further interfere with JQB (retroactive inhibition). These assumptions suggest that as time passes, $S$ may replace JQB with a more common sequence—perhaps JOB.

The "unit" in unit-sequence interferences is usually the word, although it could be any sequence of letters presented as an independent unit. Word-sequence interferences result when a sequence presented for learning conflicts with syntactic habits. Suppose a list presented to $S$ contains the sequence STAND DID LINGERIE THEY IN. Assumptions analogous to those given in the preceding paragraph predict that as time passes, $S$ may replace such a sequence with a more familiar one—perhaps THEY DID STAND IN LINGERIE.

To collect evidence for these two sources of interference, Underwood and Postman used serial anticipation learning and measured retention after 1 week for four kinds of lists: common and uncommon words, and common and uncommon trigrams. As time passed, errors due to remembering an item in an incorrect order increased faster for word lists than for trigram lists. In a similar experiment, Postman (1961) also found more such

intralist errors in recalling a list of common nouns than in recalling uncommon nouns. Both findings may be interpreted as evidence for word-sequence interferences. Underwood and Postman (1960) also found that as time passed, letter-sequence errors increased faster in the uncommon trigram lists than in the common trigram lists. This may be interpreted as evidence for letter-sequence interferences.

On the other hand, when the criterion is number of items recalled correctly, neither of the interferences have been revealed unless the analysis was restricted to the middle serial positions—to items that had not been overlearned. Thus evidence for the effect of these two sources of extra-experimental interference is not as strong as one would wish.

The most direct and convincing evidence for sequential interferences would be to demonstrate that uncommon sequences are replaced by more common ones during recall. This investigation will attempt to demonstrate this by a variation of serial reproductions which might be called serial reconstructions.

## PROCEDURE

The Ss were 27 introductory psychology students from the Johns Hopkins University. *Experiment I.*—The first S was presented with a list of letters and asked to memorize its order. Then he was given all the letters—each one typed on a separate card—and was asked to arrange the cards in the correct order. His ordering was typed and presented to the second S to memorize and to arrange in order, and this second ordering was typed and presented to the third S, and so on.

Eleven successive Ss were instructed to memorize and reconstruct the order of six lists of scrambled letters. The Ss were tested individually, one after the other, each one trying to memorize the six orderings of the preceding S. The experiment was discontinued after the eleventh ordering because

the orderings appeared to have reached a limit of no further approach toward English.

Six sentences ranging from 25 to 33 letters in length (counting spaces as letters) were selected from a second grade reader, and the individual letters in each sentence were scrambled. For instance, HE SAT DOWN might yield SHT EAW NDO.

The first S was given 20 sec. to study each list (by complete presentation). Then he was given all the original letters of the list—each letter typed on a separate card—and was asked to arrange them in the correct order. Before he began to arrange them, he was required to write down all he could recall. Because arranging the cards sometimes took him as long as 10 min., as he arranged the cards, he refreshed his memory by referring to these notes if he liked. The notes were discarded, but his ordering of the letters was typed and presented to the second S to memorize. The second S was given the same cards (after they had been shuffled) and asked to arrange them in order, and this second ordering was typed and presented to the third S, and so on.

With each new ordering, the list gradually approached the order of English spelling, so the time given to study a list was gradually reduced. As far as possible, E adjusted the time so that S made from three to six errors when ordering a list. By the eleventh ordering, the time had been reduced to 12 sec. for a list.

To reduce the serial position effect, the Ss were told: "Most people remember the first and last portions of a list best. Since you must arrange all the cards, you will do better if you spend proportionately more of your time memorizing the middle of the list."

*Experiment II.*—For word sequences, Exp. I was repeated with the following changes: (a) For materials, six sentences ranging from 18 to 24 words in length were selected from "Storyville Days and Nights" by Louis Armstrong (1954) and the words in each sentence were scrambled. (b) The first S was given 35 sec. to memorize a list, and by the final ordering this had been reduced to 13 sec. (c) Sixteen Ss (or 16 orderings) were needed before the lists reached a limit of no further approach toward English.

## RESULTS

*Letter-sequence interferences.*—Experiment I clearly showed that Ss replaced uncommon letter sequences with more common ones. Although

## TABLE 1

ORIGINAL LIST AND ELEVENTH ORDERING FOR FIVE LISTS OF SCRAMBLED LETTERS

| Original List | Eleventh Ordering |
|---|---|
| (a) NEHSMFECR DI VARWER HOTIHTHF ER → | MERCHIV REAHER SFDT THON HFWI |
| (b) HNEWTWY NHETO NT HTKN AOSNDO → | SWYTHE TKNOT ANDNO WEHN TOHN |
| (c) WFD ALHOD HONIO NOOEK WDE ETTU → | ALH KEETE WOUD HUOTN DOWD FOOI |
| (d) TEFTLH EAGN HGRE PULA EHT → | TEATHELD TUG NAHL EFH ERG |
| (e) EHH NC TEHEHM IRATDL IRWCHDE → | CHH CHEEME HUDWAR DIRLT HEIT |

they were asked to reconstruct the presented order exactly, with the first four or five reconstructions at least, the lists came closer and closer to English spelling.

Here are several orderings for a list selected at random:

Original    IE TYOUW LHRHES MNHKWEIH CNE ETAR

First    AI THYHLE MUHK WEIO EHRRWN ETCESN

Second    AI HYTWE WENKO HICH ERRHN MELTESN

Sixth    AI WEKNO HICH NERYHUE SWERT METHL

Eleventh    OWI HWE HEUI ACKRE NTHSRN MEYTHEL

The original and the eleventh orderings for the other five lists are given in Table 1.

It may not be obvious that the final orderings in Table 1 resemble English



FIG. 1. Mean trigram frequency for six lists of letters arranged in order of successive orderings by Ss.

more than the originals, but Fig. 1 shows quantitatively that Ss did replace infrequent sequences with ones that more closely resembled English spelling. In Fig. 1, mean frequency per trigram for all six lists is plotted against successive orderings. For every list, all the three-letter sequences were written down. Spaces were not considered characters. (For instance, IE TYOUW LHRH yields TYO, YOU, OUW, LHR, HRH.) The frequency with which each trigram occurred in 141,000[2] English words was determined (Underwood & Schulz, 1960, Total Count); and mean frequency was plotted against successive orderings. For the first four or five orderings at least, it is apparent that Ss replaced presented sequences with sequences that more closely resembled English spelling. Tau between frequency and ordering is .36 which is significant beyond the .05 level. Note that 6 of the 11 Ss were tested at the horizontal part of the curve; therefore this tau is a very conserva-

[2] Underwood and Schulz (1960, p. 75) state that their Total Count represents frequencies of occurrence in 1,035,000 words. However, B. Underwood (personal communication) has indicated that this is in error. In Appendix D, the Thorndike-Lorge (1944) count was based on a sample of 2,080 of the 19,440 words in the The Teacher's Word Book of 30,000 Words, and so represents a base closer to 106,000 than 1,000,000 words. Thus, the Underwood and Schulz Total Count represents frequencies of occurrence in approximately 141,000 words.

TABLE 2

Original List and Sixteenth Ordering for Five Lists of Scrambled Sentences

| Original List | Sixteenth Ordering |
|---|---|
| (a) ABOUT WAS GOOD-LOOKING WAY AND TREATING MADE OF THAT A HIM THE QUIET YOUNGSTER NICE HE MANNERS A THEM GIRLS WILD GO WITH | HE WAS A YOUNGSTER NICE QUIET WITH MANNERS GOOD-LOOKING AND A WAY OF TREATING THEM THAT MADE THE GIRLS GO WILD ABOUT HIM |
| (b) AS BE CHILDHOOD TO LIVED FROM FRIENDS BEEN CONTINUE REAL HAD TRUE AS WE LONG WE AND WOULD WE | WE LIVED AS FRIENDS LONG TRUE FROM CHILDHOOD AND WE WOULD BE REAL TO CONTINUE AS WE HAD BEEN |
| (c) I BACK AND LONG AND SITTING ALL WORK DAY MULE MY MY IT SHOVELING GET BEHIND TO AWFUL USED PAINS COAL IN HARD WAS | I AND MY ALL DAY LONG SHOVELING COAL IT WAS HARD AWFUL WORK AND SITTING BEHIND MULE USED TO GET PAINS IN MY BACK |
| (d) GOOD WE BIG AND ME A COOKED IRENE GUMBO SEE OF HIM POT FOR HE AND CAME TO | GOOD CAME ME AND FOR HIM IRENE COOKED A BIG POT OF GUMBO TO SEE HE AND WE |
| (e) THEM HEART I WORLD OF MY FROM THE TO THE BOTTOM BE REPLACE WILL ABLE NEVER THAT SAY AND | NEVER SAY THAT WORLD AND BE MY HEART ABLE TO REPLACE THE I WILL FROM THE BOTTOM OF THEM |

tive estimate of the correlation between frequency and ordering.[3]

Figure 1 shows that after the fifth ordering—after the sequences reached a mean occurrence of about 130 occurrences per 141,000 words—there was little if any further approach toward English spelling. Although the orderings changed somewhat, they remained at about the same level of nonsense. This suggests that the Ss deliberately tried to reproduce the same level of nonsense even when they had forgotten the exact sequence.

*Word-sequence interferences.*—Experiment II repeated the above results for word sequences: the Ss replaced uncommon word sequences with more common ones. With each new reconstruction, the arrangement came closer and closer to a sensible English sentence.

[3] These reconstructions were also scored as to bigram frequency per 121,000 words. The curve was almost identical to Fig. 1. It reached a peak at the fourth reconstruction, fell slightly, and began to rise again at the ninth reconstruction.

Several orderings for a list selected for its inherent interest are given:

(Original) STAND DID LINGERIE THEY IN STORYVILLE THE DID AS NOT NEIGHBORHOOD FINE IN DOORWAYS SILK THEIR WEARING GIRLS OUR IN. (First) STAND THEY DID IN STORYVILLE LINGERIE FINE THEIR THE DID NOT IN SILK NEIGHBORHOOD AS DOORWAYS WEARING GIRLS OUR IN. (Second) STAND DID THEY STORYVILLE LINGERIE FINE THEIR IN SILK NEIGHBORHOOD OUR THE NOT GIRLS DID WEARING IN DOORWAYS IN AS. (Fourth) FINE DID THEY STAND IN STORYVILLE LINGERIE IN NEIGHBORHOOD SILK OUR NOT GIRLS THE WEARING DID IN DOORWAYS AS THEIR. (Seventh) AS THEY DID STAND STORYVILLE LINGERIE IN SILK NEIGHBORHOOD IN OUR DOORWAYS THEIR NOT WEARING GIRLS FINE DID IN THE. (Eleventh) AS THEY DID NOT STAND STORYVILLE LINGERIE IN THEIR FINE SILK DOORWAYS IN OUR NEIGHBORHOOD DID THE GIRLS WEARING IN. (Sixteenth) THEY DID NOT STAND IN THEIR FINE SILK STORYVILLE LINGERIE IN THE DOORWAYS WEARING AS OUR GIRLS DID IN NEIGHBORHOOD.

In Table 2, the original and the sixteenth orderings are given for the other five lists. Even if it is not clear from the examples that the sixteenth
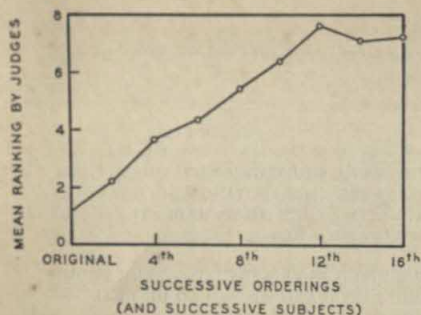
FIG. 2. Mean ranking by judges for six scrambled sentences arranged in order of successive orderings by Ss.

orderings resemble English more than the originals, Fig. 2 shows quantitatively that Ss did replace unfamiliar word sequences with more familiar ones. In Fig. 2 successive orderings are plotted against mean rank as to sensibleness. For each scrambled sentence, the original and eight even orderings were typed on individual cards, shuffled, and 10 judges were asked to rank them as to their approximation to English.[4] Tau between ordering and mean rank was .89, which is significant beyond .0005.

Figure 2 suggests that the orderings reached a limit of no further approach toward English somewhat before the sixteenth ordering, and the examples show that the sixteenth ordering is still far from a sensible English sentence.

### DISCUSSION

The fact that Ss replaced uncommon sequences with more common ones is a rather direct demonstration of the extra-experimental interference from language habits, and further discussion would add little of value; however, three procedural

[4] The judges were all graduate students familiar with approximations to English. Six coefficients of concordance were computed—one for their orderings of each sentence. Mean coefficient of concordance was .88, significant beyond the .001 level. Clearly the judges agreed among themselves.

details are worth examining. They do not raise serious doubts about the main conclusions, but they may influence the shape of the functions in Fig. 1 and 2.

First we should examine the consequences of E's attempt to hold errors constant by progressively reducing study time as the lists became progressively more simple. For the later Ss, this constant-error criterion forced errors on sequences that were almost correct English. Thus it raised the asymptote and reduced the negative acceleration somewhat. The more usual constant time limit produces a sharper negative acceleration: as soon as the list comes close to English, Ss are able to memorize it completely and the curve levels off.

Second, each S was exposed to and tested upon six successive lists totaling over 130 successive items. On the later lists, he was clearly subject to considerable proactive interference from the earlier sequences and from his own responses on earlier tests. Proactive interference from S's earlier lists should have no effect at all if we assume that there are not extraexperimental interferences from language habits. Interference from lists would cause him to replace one trigram (for instance) with another from an earlier list. Some replacements would be of more common occurrence in English, some of less common occurrence. If the replacements were consistently of more common occurrence, then this would argue for extraexperimental interference from language habits. Proactive interference from earlier tests would lead S to replace a trigram with a trigram from his own earlier reconstructions. This would raise mean trigram frequency because the mean trigram frequency of the reconstructions tended to be higher than the lists. Clearly this is a second-stage effect of extraexperimental interference from language habits.

Third, it should be emphasized that the test was not pure reconstruction: it was reconstruction confounded with recall because Ss were required to jot down their recall of the list and were allowed to consult these notes during reconstruction. The reconstruction tech-

nique was simply a stratagem to prevent the loss of elements that occurs in the usual serial reproduction technique.

In addition to the three procedural points, a marginal finding is worth examining—the finding that long before the reconstructions became sensible English, they reached a limit of no further approach toward English. After this point, although the orderings changed somewhat, they remained at about the same level of nonsense. Apparently Ss were deliberately trying to reproduce the same level of nonsense.

In fact, several Ss reconstructing scrambled letters commented that when they were unsure of a sequence, they would be certain that it was not a real word. For if it had been a real word, they said, they were sure they would have remembered it. Therefore in such a situation, they would purposely arrange the letters so that they did not spell a word.

To check the fact of the asymptote, two lists of each kind were reconstructed 10 more times by another group of Ss. The letter lists dropped from a mean of 128 occurrences per 141,000 words at the eleventh reconstruction to 124 at the twenty-first. In short the fourth reconstruction resembled English spelling about as closely as the twenty-first reconstruction.

## SUMMARY

Two extraexperimental sources of interference were examined: (a) letter-sequence interferences, which result when a list presented for learning conflicts with English spelling habits, and (b) word-sequence interferences, which result when a list presented for learning conflicts with the habits of English syntax. Both interferences were studied by serial reconstructions. The first S was exposed to a list and then he tried to reconstruct its order. Each succeeding S tried to reconstruct the order given by the preceding S. With each successive reconstruction, scrambled lists came closer and closer to sensible English. However, the orderings reached an asymptote after which there was no further approach toward English.

## REFERENCES

ARMSTRONG, L. Storyville days and nights. In, *New world writing*. New York: Mentor, 1954.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, 47, 285–293.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, 61, 97–110.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73–95.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

# THE EFFECT OF ORDER OF APPROXIMATION TO THE STATISTICAL STRUCTURE OF ENGLISH ON THE EMISSION OF VERBAL RESPONSES [1]

KURT SALZINGER, STEPHANIE PORTNOY, AND RICHARD S. FELDMAN

*Biometrics Research, New York State Department of Mental Hygiene and Columbia University*

The variables controlling the emission of verbal behavior may be roughly divided into response-produced stimuli (the speaker's verbal responses surrounding the "controlled" response) and external stimuli (audience, reinforcements, etc.). The present study deals with the effect of response-produced stimuli as specified by Miller and Selfridge's (1950) passages of different orders of approximation to the statistical structure of English.

These passages were originally constructed for the purpose of studying the effect of sequential association upon memory. It was hypothesized that the better memory for higher orders of approximation to English is related to the greater number of words determining each subsequent word. Many investigations have made use of these or similarly constructed passages as independent variables for such factors as memory (Deese & Kaufman, 1957; Marks &

Jack, 1952; Miller & Selfridge, 1950; Richardson & Voss, 1960; Sharp, 1958), shadowing of one of two dichotic messages (Moray & Taylor, 1958), "meaningfulness" (Marks & Taylor, 1954), and eye-voice span (Lawson, 1961). Therefore it was thought worthwhile to obtain an exact measure of the amount of contextual dependency in each passage in order to test the assumption that consecutive orders are equally distant from each other. Obviously, if they are not, the shape of any curves employing this assumption in relating the above mentioned factors to approximation to English would be inaccurate.

The technique chosen for evaluating the amount of contextual dependency was the "cloze" procedure (Taylor, 1953, 1954, 1956), which was originated as a measure of readability and which consists of having Ss guess the words which have been systematically deleted from a given passage. Thus, application of the cloze procedure to these passages will—in addition to supplying an exact measure of contextual dependency—provide standards for comparing different texts, e.g., one text might be described to be as readable as third-order, another as readable as seventh-order approximation to English.

Finally, this study will provide the opportunity to examine the effect of order of approximation upon the number of guessed words which fall into the same grammatical category

of speech as the "correct" word. This will make it possible to evaluate the contribution which the syntax, resulting from these approximations to English, makes to such functions as memory.

## METHOD

*Passages.*—The 50-word passages of Miller and Selfridge (1950), at each of the eight orders of statistical approximation to English (zero, first, second, third, fourth, fifth, seventh, and Text), were the basic material for this study. Each passage was mimeographed on a separate sheet of paper, each deleted word being indicated by an underlined blank of constant size. The order of presentation of the passages was randomized within each set of eight and varied from $S$ to $S$.

*Experimental groups.*—Ninety-three undergraduate students, with a median age of 18.9 yr., were required to participate in this experiment as part of their regular course work.

Group A₁ (17 $S$s, 11 male and 6 female) and Group A₂ (16 male $S$s) were given the passages with every fifth word deleted, starting with the fifth word of each passage and omitting the last word. Groups A₁ and A₂ gave essentially the same results on all measures, and therefore they have been combined into a single group (A₁A₂) in some of the analyses presented below. For all analyses of grammatical category, only Group A₂ is included since it was assumed that further analysis of Group A₁ would have yielded the same results.

Group B (33 male $S$s) was given the same passages as Groups A₁ and A₂, with every fifth word deleted, starting with the sixth word of each passage and omitting the first word.

Group C (27 male $S$s) was given the same passages with every seventh word deleted, starting with the seventh word of each passage and omitting the last two words.

*Procedure.*—The passages were administered to $S$s in a group. They were told in the instructions read to them that the typescripts contained no punctuation and they themselves would have to decide where a new thought began, that only one word was called for in each blank, and that, when they were not certain, they should guess rather than skip a blank. Thirty minutes were allowed for completion of the task.

## RESULTS AND DISCUSSION

*Proportion of words correctly guessed.*
—Figure 1 shows that the proportion
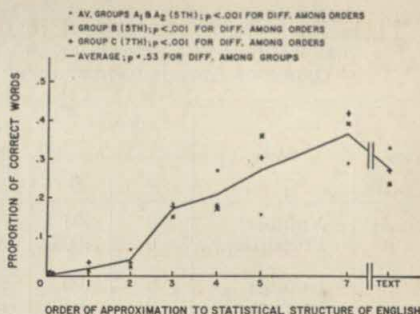


FIG. 1. Proportion of words guessed correctly as a function of order of approximation to English.

of correctly guessed words increased from Orders 0 through 7, with Text approximately equal to Order 5. Friedman analyses of variance (Siegel, 1956), used for this and all subsequent comparisons, showed that the three groups (A₁A₂, B, C) did not differ significantly from one another ($P = .53$) and that, within each group, the increase over orders was statistically significant ($P < .001$ for each group). Therefore, the mean curve in Fig. 1, based on $N = 93$ and on about 40% of the words of each passage, represents a more reliable summary of these data than each curve separately.

This result also makes clear that having 6 words on either side of the blank does not produce more correct guesses than having 4 words on either side of the blank, thus reinforcing the conclusion of Taylor (1954) and MacGinitie (1961) that the words predicted for every fifth blank are independent of each other, and the conclusion of Aborn, Rubenstein, and Sterling (1959) that 11-word sentences result in optimum predictability in comparison to 6-word and 25-word sentences. Apparently $S$s either do not or cannot make use of a context of more than 5 words on either side of each blank.

The fact that the relationship between order and proportion of correct

TABLE 1

MEAN NUMBER AND PROPORTION OF EXACT CORRECT WORDS AS A FUNCTION OF
ORDER OF APPROXIMATION TO THE STATISTICAL STRUCTURE OF ENGLISH

| Group | Score | Order of Approximation | | | | | | | Text |
| | | 0 | 1 | 2 | 3 | 4 | 5 | 7 | |
|---|---|---|---|---|---|---|---|---|---|
| $A_1A_2$ | Number[a] | 0 | .03 | .61 | 1.64 | 2.45 | 1.42 | 2.61 | 2.97 |
| | Proportion[b] | 0 | .003 | .068 | .182 | .272 | .158 | .290 | .330 |
| B | Number | 0 | .09 | .21 | 1.39 | 1.58 | 3.24 | 3.55 | 2.15 |
| | Proportion | 0 | .010 | .023 | .154 | .176 | .360 | .394 | .239 |
| C | Number | 0 | .29 | .29 | 1.61 | 1.61 | 2.72 | 3.78 | 2.45 |
| | Proportion | 0 | .032 | .032 | .179 | .179 | .302 | .420 | .272 |
| Mean Number | | 0 | .14 | .37 | 1.55 | 1.88 | 2.46 | 3.31 | 2.52 |
| Mean Proportion | | 0 | .016 | .041 | .172 | .209 | .273 | .368 | 2.80 |

[a] Number of correct words per $S$.
[b] Proportion of correct words to total number of words guessed.

words is not linear means that the assumption of equal distances between consecutive statistical orders is not tenable. If we utilize the values in Table 1 as correct estimates of the distances between orders of approximation, then the distance between Orders 2 and 3 is more than five times the distance between Orders 1 and 2. For memory improvement on these same passages[2] (Selfridge, 1949), the corresponding ratio is less than 1.5. Figure 2 shows the memory data plotted in two ways: (a) according to the assumption of equal intervals between orders of approximation, and (b) with orders spaced according to number of words correctly guessed in the present study. From Orders 3 through 7, the curves are essentially the same. Up to Order 3, however, the intervals between successive orders vary greatly in size, indicating that

[2] The set of passages used in the present study is available in Miller and Selfridge (1950). However, their memory data are based on averages of two passages at each order of approximation. The percentages of words correctly recalled, for the single set of passages used here, were obtained from Selfridge's (1949) thesis, and correspond to her first set of lists.

the equal-intervals assumption is least tenable in this region. One additional fact is that Text gives rise to fewer correct words than Order 7. This indicates that it was misplaced on the order continuum and also suggests that different texts might well appear in different positions on this continuum.

*Proportion of different words.*—Figure 3 shows a plot of the proportion of the total number of guessed words which were different. With the exception of Text, which is again out of line, it is clear from this graph that increasing order of approximation not only increases the probability of evoking a correct response but also limits the variety of different responses. The difference among the three groups of $S$s approaches significance ($.05 < P < .08$), but the graphs of the separate groups each show the general tendency to decrease over orders. The decrease is significant in each case ($P < .008$ for $A_1A_2$, $P < .001$ for B, $P = .002$ for C).

Figure 4 shows a plot of the proportion of the total number of different *incorrect* words. With the exception of Text, there is a general decrease in the proportions with increasing

order, although the overall drop is only about half of what it is in Fig. 3. The differences among the three groups approach significance (.05 < P < .08) as before, but here this is due to the fact that Group C does not show a significant decrease over orders ($P = .20$) while Groups $A_1A_2$ and B do show significant decreases ($P < .05$ and $P < .04$, respectively). This is the only measure on which one of the groups differed significantly from the others. A suggested explanation is that, with only six blanks for each passage in



FIG. 3. Number of different words including the correct word (expressed as a proportion of the total number of words guessed) as a function of order of approximation to English.



ORDER OF APPROXIMATION TO STATISTICAL STRUCTURE OF ENGLISH

FIG. 2. Percentage of words correctly recalled (Selfridge, 1949) as a function of order of approximation to English, with an equal-interval scale for orders (dashed line), and with the intervals determined by the average number of words guessed correctly in the present study (solid line).

which every seventh word was deleted, the total number of guessed words may not have been large enough to allow significant changes in number of different words from order to order when only incorrect guesses are considered.

*Grammatical classification.*—The grammatical classification was based on Fries' (1952) system of analysis. The two groups of words called *general* grammatical categories below are lexical words (roughly equivalent to the classical categories of nouns, pronouns, verbs, adjectives, and ad-

verbs) and function words (roughly equivalent to articles, conjunctions, prepositions, auxiliary verbs, interjections, and quantity words). The five *specific* grammatical categories are nouns and pronouns considered as one class, verbs, adjectives, adverbs, and function words as already described.

Figure 5 shows that, as order of approximation increases from Orders 0 to 7, an increasing proportion of the words emitted by Ss belongs to the same specific grammatical category as the deleted words. The increase is significant for each group ($P < .001$ for $A_2$, B, and C), and the groups do not differ from one another ($P = .24$).



ORDER OF APPROXIMATION TO STATISTICAL STRUCTURE OF ENGLISH
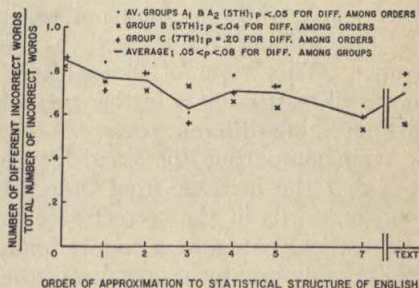
FIG. 4. Number of different incorrect words (expressed as a proportion of the total number of incorrect words) as a function of order of approximation to English.
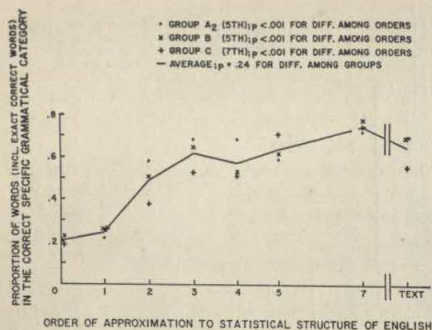
FIG. 5. Proportion of guessed words belonging to the same *specific* grammatical category as the correct word, as a function of order of approximation to English.

This graph again shows performance on Text to be worse than at Order 7 and only slightly better than at Order 5. It should be noted that at Order 0 *S*s perform very close to chance level, i.e., when the emitted words are put into the five specific grammatical categories, about one-fifth fall into the correct category.

Figure 5 differs from Fig. 1 (number of correct words) in that it seems to approach an asymptote at about Order 3 with only a relatively small increase after that, while the number of correct words rises as much after as before Order 3. It is also apparent that the maximum proportion of words in the correct grammatical category is more than twice as large as the proportion of exact correct words, i.e., while *S*s may not react to the meaning they react to the syntax. This type of "knowledge" may well be used as a further way of distinguishing different texts.

Again comparing Fig. 5 and 1, it is seen that the increase from Orders 1 to 2 for words in the correct specific grammatical category accounts for a considerably greater proportion of the overall increase than does the corresponding change for exact correct words. Beyond Order 3, as already stated, correct specific grammatical

classifications appear to approach an asymptote while number of exact correct words continues to increase. Therefore, if grammatical classification of guessed words is taken as a measure of amount of syntactical structure, and exact correct words are taken as a measure of amount of meaning, improvement in memory must be attributed primarily to increased syntactical structure between Orders 1 and 2, about equally to syntactical structure and to meaning between Orders 2 and 3, and primarily to meaning beyond Order 3.

If we examine the relationship between order and specific grammatical category, taking into account only those responses which differed from the deleted words, it is found to be nearly identical to that shown in Fig. 5, where correct responses are included. The increases over orders are significant ($P < .001$ for $A_2$, B, and C) and the three groups do not differ from one another ($P > .15$).

The relationship between order and general grammatical category also remains nearly the same whether or not correct responses are taken into consideration. Again at Order 0 *S*s perform very close to chance level, i.e., when the emitted words are put into the two general grammatical categories of lexical and function words, without further subdivision, about one-half fall into the correct category. In general these relationships are very similar to the analogous ones for specific grammatical categories. Although the increases over orders cover a smaller range when the words are divided into only two categories, they are significant for all three groups (including correct responses: $P < .002$ for $A_2$, $P < .001$ for B and C; not including correct responses: $P < .02$ for $A_2$, $P < .001$ for B and C), and the groups do not

differ from one another ($P = .53$ in both cases).

These results are in general agreement with Epstein's (1961) finding that learning of a syntactically structured series of words is superior to that of a random, unstructured series of words, since the two series correspond to Order 1 and Text of the present study. Epstein found in addition that there was no significant difference between his *Ss'* ability to learn syntactically structured nonsense syllable material and their ability to learn syntactically unstructured meaningful material, and this result demonstrates that syntax or meaning alone contributes to learning. This makes more plausible our suggestion that both "meaning" and syntax make contributions to memory for Miller and Selfridge's (1950) passages.

## SUMMARY

Ninety-three undergraduate students were required to guess the words that were systematically deleted from a series of passages varying in order of approximation to the statistical structure of English. The *Ss* guessed a greater proportion of words the higher the order of approximation to English. Proportion of words in the correct grammatical category also showed an increase with increasing order of approximation. Proportion of words in the correct grammatical category increased most from Order 0 to Order 3 while proportion of correct words continues to increase as much after Order 3 as before it. Thus analysis of the predicted words into their grammatical categories may well be useful for further differentiation of text materials, since the two types of analysis do not yield the same type of information. Finally, it was pointed out that the assumption of equal intervals between successive orders of approximation is untenable, and the relationship between memory and order of approximation to English can be explained in part by syntactical structure and in part by the "meaning" called for by the context.

## REFERENCES

ABORN, M., RUBENSTEIN, H., & STERLING, T. D. Sources of contextual constraint upon words in sentences. *J. exp. Psychol.*, 1959, 57, 171–180.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, 54, 180–187.

EPSTEIN, W. The influence of syntactical structure on learning. *Amer. J. Psychol.*, 1961, 74, 80–85.

FRIES, C. C. *The structure of English.* New York: Harcourt, Brace, 1952.

LAWSON, E. A. A note on the influence of different orders of approximation to the English language upon eye-voice span. *Quart. J. exp. Psychol.*, 1961, 13, 53–55.

MACGINITIE, W. H. Contextual constraint in English prose paragraphs. *J. Psychol.*, 1961, 51, 121–130.

MARKS, M. R., & JACK, O. Verbal context and memory span for meaningful material. *Amer. J. Psychol.*, 1952, 65, 298–300.

MARKS, M. R., & TAYLOR, W. L. The influence of contextual and goal constraints on the meaningfulness of "automatic sentences." *J. soc. Psychol.*, 1954, 40, 43–51.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, 63, 176–185.

MORAY, N., & TAYLOR, A. The effect of redundancy in shadowing one of two dichotic messages. *Lang. Speech*, 1958, 1, 102–109.

RICHARDSON, P., & VOSS, J. F. Replication report: Verbal context and the recall of meaningful material. *J. exp. Psychol.*, 1960, 60, 417–418.

SELFRIDGE, J. A. Investigations into the structure of verbal context. Unpublished honor's thesis, Harvard University, 1949.

SHARP, H. C. Effect of contextual constraint upon recall of verbal passages. *Amer. J. Psychol.*, 1958, 71, 568–572.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TAYLOR, W. L. Cloze procedure: A new tool for measuring readability. *Journalism Quart.*, 1953, 30, 415–433.

TAYLOR, W. L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Michigan, 1954.

TAYLOR, W. L. Recent developments in the use of cloze procedure. *Journalism Quart.*, 1956, 33, 42–48.

# THE EFFECTS OF REWARD AND KNOWLEDGE OF RESULTS ON THE PERFORMANCE OF A SIMPLE VIGILANCE TASK

RAYMOND R. SIPOWICZ, J. ROGER WARE, AND ROBERT A. BAKER

*United States Army Armor Human Research Unit, Fort Knox, Kentucky*

Although the effects of reward and knowledge of results on human performance have been intensely studied, the effectiveness of such variables on human monitoring has received relatively little attention. In view of the practical importance of vigilance research and the need to sustain and maintain a high level of detection, this neglect is somewhat surprising. Pollack and Knaff (1958) studied the effects of reward and punishment on monitoring, and reported that punishment was more effective than reward. Moreover, the reward condition proved relatively ineffective in improving performance beyond a level obtained without reward.

Mackworth (1950); McCormack (1959); Loeb and Schmidt (1960); and Weidenfeller, Baker, and Ware (1962) did, however, obtain significant improvement in vigilance performance using knowledge of results (KR). In the last study, not only KR but "false KR" was also found to be effective. Knowledge of results in this experiment was given by means of a bright white light. While it might be assumed that such added stimulation would raise *S*'s activation level and hence increase detection probability, the inclusion of a control group having the light alone (administered in synchrony with the typical vigilance decrement) showed no arousal effects. A similar finding has been reported by Davis, McCourt, and Solomon (1960).

The effectiveness of KR in these studies, and the apparent failure of

"reward" in the Pollack and Knaff (1958) study, is somewhat contradictory—especially in view of the fact that feedback or KR can be, and often is, logically regarded as a form of reward or reinforcement. In the Pollack and Knaff study no mention was made of the amount of reward except to say that it was an extra hour of pay. Further, the conditions of reward were such as to make it difficult for more than 1 or 2 *S*s in the group to be rewarded.

If, however, in the typical vigilance study the reward is contingent upon the maintenance of a certain level of vigilance by each *S*, the effects of the reward might be more pronounced than in a situation where a reward is given for each signal detected or on a competitive basis where only the best are rewarded. Moreover, if reward and KR are individually effective, a condition combining the two might be doubly effective.

In order to furnish answers to some of these questions, the effects of KR, reward, and the combination of the two in a typical vigilance task were studied under the conditions in which a monetary reward was individually administered and was contingent upon the maintenance of a high level of performance.

## METHOD

*Subjects.*—Eighty Fort Knox armor trainees, aged 17 to 24 yr., free from visual defects, served as *S*s. Twenty *S*s were arbitrarily assigned to Group C (Control), 20 to Group KR, 20 to Group R (Reward), and 20 to a

combined KR and reward group (Group R + KR).

*Apparatus.*—The Ss' task was to detect aperiodic interruptions of a continuous light source over a 3-hr. period. The light source (12 v. dc bulb, .45 ft-c as a point source, operating at 5½ v.) was located at approximately eye level in a flat black plywood box. The schedule of interruptions (signals) was predetermined on the basis of 12 signals per ½ hr. with a total of 72 for the 3-hr. session. The randomized intersignal intervals ranged from 24 to 360 sec. with an average interval of 150 sec. The schedule was presented by means of a film-tape, fed through a Gerbrands variable-interval programer, and a simple timing circuit. The latter determined the duration of the interruption (.03 sec. measured at the timing relay with a Hunter Klockounter), and the former determined the intervals between the interruptions. The signal presentations and Ss' signal detections were recorded by a 20-pen Esterline-Angus operations recorder.

*Procedure.*—Each S, wearing earphones to reduce ambient noise, monitored the light in an isolated room. For the no-KR condition, S was told to press the response button as soon as he saw a signal. Each S was further told that the signals could occur at any time, so it was necessary to remain alert and watch the light at all times. To insure that S understood the requirements, a practice period of 10 signals (at 1-min. intervals) was given before the watch session began.

For the KR conditions, a 1½-in. pilot lamp was installed at approximately eye level on the right side of the monitoring display. The S's requirements were the same as before; S was told that the pilot lamp would light if he missed a signal, thus indicating a signal was missed and that he should be more alert in order to detect subsequent signals. This light, unlike the signal light, was covered by a white filter lens and operated at its maximum voltage—6 v. dc. The light was flashed for a 2-sec. period if a given signal was not detected within 5 sec. after its presentation.

For the reward conditions, each S was told he would receive $3.00 if he detected all the signals. In addition, for each signal missed S was told that he would lose a portion of the initial $3.00 in geometric progression beginning with 5 cents, for the first miss. By missing six or more signals during the watch session, S lost the entire amount. Thus, if S missed only one signal in the 3-hr. period he received $2.95; for two misses, he received $2.85; for three, $2.65; for four, $2.25; for five, $1.45, and with six or more misses, nothing. As in the KR condition, a miss was scored if S failed to report a signal within 5 sec. after its appearance.

Although the study was designed for 20 Ss per group, several Ss fell asleep during the watch session. Since Ss could not be directly observed, it was arbitrarily decided to eliminate from the study all Ss who missed 12 or more consecutive signals. This was done on the assumption that Ss missing 12 or more consecutive signals were either asleep or uncooperative. On this basis, 5 of the original 80 Ss were rejected—2 from Group C, 1 from Group KR, and 2 from Group R. None of the 20 Ss, however, were eliminated from Group R + KR. In an attempt to equalize the number of Ss in each group, additional sessions were held. Due to experimental difficulties, however, data were obtained on only 19 Ss in Group KR.

## RESULTS

The mean percentage of signals missed by each group, by 30-min. periods, is shown in Table 1. On an

TABLE 1

MEAN PERCENTAGES OF SIGNALS MISSED FOR ALL GROUPS OVER SIX 30-MIN. PERIODS AND OVER ENTIRE 3-HR. SESSION

| Group | Successive 30-Min. Periods | | | | | | | | | | | | | Overall | |
| | 7 | | 2 | | 3 | | 4 | | 5 | | 6 | | | | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 18.2 | 12.6 | 21.6 | 16.5 | 19.2 | 13.6 | 25.0 | 15.1 | 27.4 | 17.4 | 34.1 | 27.5 | 24.3 | 12.1 |
| KR | 7 | 8.1 | 10.8 | 11.1 | 12.2 | 11.3 | 14.4 | 11.5 | 14.4 | 19.1 | 12.3 | 17.7 | 11.8 | 10.5 |
| R | 8.7 | 11.9 | 5.9 | 7.7 | 5.2 | 9.5 | 7.9 | 8.4 | 11.6 | 13.5 | 10.5 | 11.2 | 8.3 | 7.1 |
| R + KR | 4.4 | 4.9 | 2.0 | 3.5 | 6.2 | 7.1 | 5.8 | 6.7 | 3.6 | 4.9 | 3.6 | 4.9 | 4.3 | 2.4 |

overall basis, Group C missed 24.3% of the signals; Group KR, 12%; Group R, 8.4%; and Group R+KR, 4.3%. The raw data in terms of number of signals missed by the groups were analyzed to obtain means and variances. A test for homogeneity of variance was performed on these data and resulted in a significant $F$ value. The data were transformed to a log scale and then subjected to an analysis of variance.

This analysis showed the groups were significantly different ($F = 20.92$, $df = 3$, $P < .01$). Using Duncan's (1955) multiple range test, all the experimental groups were found to differ significantly from the control ($P < .01$). The mean difference between Group R + KR and Group KR was also significant at the .01 level. The difference between Group R+KR and Group R was, however, significant at only the .05 level. No significant difference was found between Group R and Group KR.

An inspection of the performance curves also revealed the typical vigilance decrement within the first hour for Group C. This expected decrement, however, was not shown by any of the experimental groups. For the experimental groups, the greatest decrement occurred for Group KR between the first and fourth 30-min. periods. A correlated $t$ test for the decrement, however, showed the difference was nonsignificant.

Although it might be argued that knowledge of results should lead to an increase in the number of false responses, an analysis of the present data showed no significant differences between groups. The $Ss$ in Group KR averaged 1.05 false responses as compared with 0.84 for Group C, 0.72 for Group R, and 0.25 for Group R+KR.

## DISCUSSION

Of most importance was the finding that the combination of reward and knowledge of results led to the detection of better than 95% of the signals—a performance at a "near perfect" level. Further, the extremely small variance of this group may be some indication of the extent to which individual differences in performance can be decreased by the use of such motivational techniques. Classification of $Ss$ into "good performers" and "bad performers" as was done in other studies (Buckner, Harabedian, & McGrath, 1960; Fraser, 1953) may be adequate in vigilance tasks in which no attempt is made to sustain a high level of motivation. Results from this study, however, suggest the performance of all $Ss$ in monitoring tasks may be sustained at a high level. Thus, the problem of selection and assignment of $Ss$ to vigilance tasks on the basis of their initial ability can be eliminated if effective motivational techniques are discovered and properly employed.

Of additional interest is the fact that reward alone proved to be only slightly more effective than KR. It should be noted, however, that in the reward situation $S$ had no way of knowing whether or not a singal was missed and, consequently, was unable to adjust effectively his monitoring behavior to meet the requirements of the task.

The relatively high level of performance, as well as the failure to find the usual vigilance decrement for the two reward conditions, focuses considerable attention upon the method used in dispensing the reward. First of all, $Ss$ did not have to detect all of the signals in order to be rewarded. The $S$, however, was required to detect at least 93% of the signals. Moreover, $Ss$ were told before the session began that they would be given the reward. Their subsequent performance determined whether or not it would be retained. This technique, it is believed, was particularly effective in sustaining a high level of motivation

over the entire watch period. When specific knowledge of the reward situation was then made available throughout the watch session, $S$'s motivation was further increased and an even higher level of performance was obtained.

In previous studies of KR such as Mackworth's (1950), $S$s were informed whenever they made a false response, when they failed to detect a given signal, and when they correctly detected a signal. It is likely, however, that much of this information is superfluous. Knowledge of results as used in this study gave $S$s information only when they failed to detect a given signal. In terms of the mean number of signals missed, this technique was as effective as those used by Mackworth. Mackworth's experimental group missed 17.3% of the signals presented in a 2-hr. session, whereas $S$s in the present study missed only 12% over a 3-hr. period.

## SUMMARY

Four groups of $S$s (20 to a group) monitored aperiodic and brief interruptions of a continuous light source under isolated conditions for a 3-hr. period. The $S$s in Group R were given $3.00 if they detected all signals presented during the watch session, but lost .05, .15, .35, .75, 1.55, or all if they missed one, two, three, four, five, or six signals. Group KR was informed of all signals missed by a bright flash of light. Group R + KR received both KR and reward according to the schedule for Group R. Group C, a control, received neither reward nor KR.

Although all experimental groups were significantly better than the control group, the combination of reward and KR produced the highest level of signal detection. The results are interpreted as indicating that either reward or KR can be effective in maintaining a high level of vigilance. The effectiveness of reward, however, is highly de-pendent upon the manner in which it is used. The effectiveness of such incentives in improving performance and reducing inter-$S$ variability also attenuated the importance previously assigned to individual differences. It is further suggested that such differences are primarily motivational and, as such, are susceptible to experimental manipulation and control.

## REFERENCES

BUCKNER, D. N., HARABEDIAN, A., & McGRATH, J. J. A study of individual differences in vigilance performance. *USA TAGO Hum. Factors Res. Br. tech. res. Rep.*, 1960, No. 2.

DAVIS, J. M., McCOURT, W. F., & SOLOMON, P. The effect of visual stimulation on hallucinations and other mental experiences during sensory deprivation. *Amer. J. Psychol.*, 1960, 116, 889–892.

DUNCAN, D. B. Multiple range and multiple $F$ tests. *Biometrics*, 1955, 11, 1–42.

FRASER, D. C. The relation of an experimental variable to performance in a prolonged visual task. *Quart. J. exp. Psychol.*, 1953, 5, 31–32.

LOEB, M., & SCHMIDT, E. A. Influence of time on task and false information on efficiency of responding to pure tones. *USA Med. Res. Lab. Rep.*, 1960, No. 426.

MACWORTH, N. H. Researches on the measurement of human performance. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, 1950, No. 268, 25–29.

McCORMACK, P. D. Performance in a vigilance task with and without knowledge of results. *Canad. J. Psychol.*, 1959, 13, 68–71.

POLLACK, I., & KNAFF, P. R. Maintenance of alertness by a loud auditory signal. *J. Acoust. Soc. Amer.*, 1958, 30, 1013–1016.

WEIDENFELLER, E. W., BAKER, R. A., & WARE, J. R. The effects of knowledge of results (true and false) on vigilance performance. *Percept. mot. Skills*, 1962, 14, 211–215.

# SOME CONDITIONS INFLUENCING THE ACQUISITION AND UTILIZATION OF CUES [1]

## LOY S. BRALEY [2]

### University of Buffalo

A paper by Bruner, Matter, and Papenek (1955) posed the question as to the variables that control the range of environmental information that an organism acquires during some behavioral sequence. This problem of the "breadth of learning" assumes that an organism is exposed to a wide range of possible cues under most circumstances, but that responses are acquired to only a portion of these with the others remaining as undifferentiated context. While it is clear that the range of cues acquired may be differentially advantageous depending on the task requirements, the focus of their paper and the present study was on the acquisition of cues that were not immediately task-relevant, but which became so at a subsequent time. The above authors demonstrated, as conjectured by Tolman (1948), a reduction in subsequent utilization of initially irrelevant cues as a product of overlearning and higher levels of motivation.

This was but a start on an empirical question. Recent reviews by Easterbrook (1959) and Kausler and Trapp (1960) on cue utilization demonstrate the continued interest in this problem. Easterbrook, concerned with the relation of emotion to behavioral organiza-

tion took the position that an increase in emotional arousal functions to reduce the utilization of cues. Consequently, degree of behavioral organization depends on task demands for "breadth" of cues. Kausler and Trapp were in substantial agreement although their focus was on the effects of motivation on the utilization of cues in incidental learning. They suggested additional relevant variables of motivation as generalized $D$ versus incentive-oriented set, spatial contiguity of relevant and irrelevant cues, and task difficulty.

The present study was designed (a) to demonstrate with human $S$s, as Bruner et al. (1955) have done with rats, that exposure to initially irrelevant cues will facilitate later learning when these cues become the relevant discriminanda, (b) to test the hypothesis that high scores on a response-inferred motivational variable (Taylor $MA$ scale) are inversely related to acquisition and utilization of such cues, and (c) to determine whether the presence of the previously correct cue from earlier learning would differentially facilitate or retard the shift to new cues as a function of low and high motive strength, respectively. Certain clinical conceptions of anxiety (Cameron, 1951; Sullivan, 1953) would predict such a perseverative tendency.

## METHOD

*Subjects.*—Eighty extreme scorers on the $MA$ scale were selected as $S$s from a sample of 525 undergraduate questionnaires obtained at the University of Buffalo and Springfield

College. The upper 20% of this sample had an *MA* score range of 20 to 40, with a mean of 25.3; the lower 20% a range of 0 to 7, with a mean of 4.59. The mean of all scores was 13.77. Twenty-four *Ss* were female, 13 of which were high *MA* scale scorers. All *Ss* were contacted and asked to participate in such a manner as to avoid any association between the *MA* scale and the experiment proper.

Because of contradictory evidence (Taylor, 1955) on the relationship of IQ to *MA* score, a short form of the Wechsler-Bellevue, Form I (Herring, 1952) was administered. The correlation between IQ and *MA* was .05. Learning rate on the experimental task and IQ revealed a nonsignificant *r* of −.22.

*Task and stimulus materials.*—The basic task was one of concept learning with simultaneous presentation of a positive and negative instance on each trial. This is similar to the Bruner et al. (1955) procedure and that of Blum and Blum (1949), and Lashley (1942) in their contributions to the continuity controversy. Each instance consisted of two small geometric figures inside a larger geometric figure, drawn on a white 4 × 6 in. card. The three figures on any card were combinations of circles, triangles, and squares each of which could be red, blue, green, or yellow. An instance then, might be a small red circle, small green triangle, within a large blue square. The stimuli for the three stages of the experimental procedure were as follows:

Stage I: All *Ss* learned a concept where the positive cue was two small like-colored figures within a large figure of a different color. Only 12 positives were used as pretesting had indicated rapid learning for this simple concept. Since only two colors appeared on these positives, 12 combinations were possible. This also permitted appropriate combinations of the three types of figures.

The same negative instances were used in all three stages so the possibility of memorization had to be obviated. The 30 negative instances were combinations of 4 values of color, 3 of shape, and 2 of size (small or large figure). In addition, 6 negatives with all figures of the same color were included to avoid concept attainment by a different-colors cue.

Stage II: While all *Ss* continued to be reinforced for the Stage I positive cue, half of the *Ss* were exposed to 10 trials of irrelevant cues. Five of these consisted of cross-hatching in the large figure and 5 consisted of the substitution of an L shape for one of the small figures. All new irrelevant cues appeared on Stage I positive instances.

Stage III: The positive cues were now either the presence of a cross-hatched large figure or an L shaped small figure. The 30 positive instances were identical to the negatives except for the addition of cross-hatching on half, or substitution of a small L shape on the remainder. No two instances that were the same in original form were opposed during exposure to *S*. Half of the *Ss* in Stage III had the Stage I positive cue present as a negative instance on 50% of the trials. For this condition a set of 20 negatives was used consisting of the last 10 Stage I positives and the last 10 of the regular negatives.

*Design and procedure.*—The experiment was a 2 × 2 × 2 factorial design with 10 *Ss* per cell. Conditions were: (a) high and low anxiety, (b) presence or absence of irrelevant cues in Stage II, and (c) presence or absence of Stage I positive cue during Stage III.

The *Ss* were randomly assigned to one of the four experimental conditions according to anxiety grouping and were run individually. The IQ measure was obtained at the conclusion of the experimental task.

The *S* faced a plywood screen with two exposure windows at approximately eye level. With the simultaneous presentation of two cards on each trial *S* was to indicate his choice of the correct card by depressing one of two response keys fixed to the table immediately below the windows. A correct choice closed a circuit and reinforcement was delivered immediately by a small white bulb recessed between and above the windows. Position of the correct card was varied according to a prearranged randomly chosen order. The *Ss* were instructed that their job was to choose the correct one of two cards and that they would have to begin by guessing. However, the reinforcing stimulus would indicate when they were correct in their choice so that eventually they were to reach a point where they could choose the correct card each time.

In Stage I, trials were continued to a criterion of 10 successive correct responses. Eight *Ss* were discarded for failure to meet criterion by Trial 70. Stage II began with completion of the criterion trials and was, to *S*, continuous with Stage I in that there was no interruption in presentation of the stimulus materials. The experimental group was exposed to the series of 10 irrelevant cues. The control *Ss* were simply continued for 10 trials beyond criterion on the Stage I cue, giving an equal number of overlearning trials to both groups. Stage III was again continuous with Stage II and where reinforcement could only be achieved by re-

## TABLE 1

Means and SDs of Transformed Trials to Criterion for Concept Learning: Stage III

| Anxiety | Positive Cue from Stage I | Irrelevant Cues Present in Stage II | | Irrelevant Cues Absent in Stage II | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| High | Present | 5.55 | 2.55 | 3.17 | 2.04 |
| | Absent | 6.35 | 2.93 | 5.68 | 2.91 |
| Low | Present | 5.45 | 1.51 | 3.93 | 1.56 |
| | Absent | 8.13 | 1.16 | 4.40 | 2.97 |

sponding to either of the previously irrelevant cues.

### RESULTS

There was no substantial difference in Stage I acquisition trials between control $Ss$ (Mean $= 10.20$, $SD = 11.60$) and those exposed to the irrelevant cues in Stage II (Mean $= 11.51$, $SD = 12.92$). While not shown, it might also be noted that neither experimental nor control $Ss$ made any errors during the 10 Stage II trials.

Table 1 presents the transformed means and $SDs$ of the eight experimental conditions. A Bartlett test revealed nonhomogeneous variances among the conditions, necessitating a square-root transformation to meet the requirements of the analysis of variance.

Table 2 indicates a significant $F$ for the Stage II irrelevant cue exposure condition. However, rather than facilitating performance in Stage III, exposure to the irrelevant cues impaired performance. It appears then that the facilitation of subsequent learning through utilization of previously irrelevant cues, as found by Bruner et al. (1955), cannot be demonstrated with human $Ss$ under these experimental conditions.

It is also clear from Table 2 that differences in the motivational condition ($MA$ scale) bear no relationship to performance in the Stage III concept learning. Further, an analysis of variance of Stage I cue learning and $MA$ score (within $MS$ larger than between $MS$ for $df = 7/72$) was not significant.

For the third condition the analysis of variance indicates that continuation of the Stage I positive cue into Stage III facilitated performance independent of $MA$ scale level. Difficulty in shifting from a successful (positively reinforced) cue to a new one while the old one is still present was no more characteristic of the high-$MA$ $Ss$ than of the low-$MA$ $Ss$.

With two cues required for concept attainment in Stage III it is of some interest to determine differential cue difficulty. Error scores were computed for the two cues in the Stage III task and an analysis of variance run on the irrelevant cue exposure condition and the type of cue (cross-hatched ground vs. L shaped figure). A significant $F$ ($F = 10.62$, $df = 1/156$) emerged for the irrelevant cue condition, confirming the analysis by trials, as well as for the type of cue ($F = 8.88$, $df = 1/156$). The cross-hatched cue covering the entire background of the

## TABLE 2

Analysis of Variance of Transformed Trials to Criterion for Concept Learning: Stage III

| Source | df | MS | F |
|---|---|---|---|
| Anxiety (A) | 1 | 1.69 | |
| Irrelevant cue Exposure (IC) | 1 | 86.15 | 14.49* |
| Stage I positive cue present (PC) | 1 | 52.17 | 8.78* |
| A × IC | 1 | 6.06 | 1.02 |
| A × PC | 1 | .31 | |
| IC × PC | 1 | .04 | |
| A × IC × PC | 1 | 19.17 | 3.22 |
| Within | 72 | 5.95 | |

* $P = .01$.

large figure was clearly an easier cue. With no procedure available for scaling the difficulty of cues it seems reasonable to assume that this was a more novel cue generating higher attention value.

## DISCUSSION

The failure of the motivational variable to predict differential acquisition and utilization of cues deserves little comment. Strictly speaking, these results are applicable only to this task, although they raise further questions about the status of the $MA$ scale as an independent selection device.

While several studies (Bruner et al., 1955; Jeeves & North, 1956; Peterson & Peterson, 1957; Weiss & Margolius, 1954) have demonstrated positive transfer of context or irrelevant stimuli, the present results suggest additional factors that demand attention. Here, exposure to initially irrelevant cues functioned so as to inhibit rather than facilitate utilization of these cues when subsequently they became the basis for correct discrimination. Berlyne (1958) and Dember (1960) have presented data showing the preference for, and attention value of, stimulus change and novelty. Luria (1957) and Sokolov (1954) have found that orienting reactions are induced by changes in the stimulus field that serve to increase the accessibility of the organism to deal with these changes. These orienting responses are maintained or extinguished as a function of the relevance of new or changed stimuli. Consequently, in the present study, despite the similarity of immediate negative reinforcement at the start of Stage III, the stimuli constituting the positive instances were novel and attention-compelling for the control $S$s while for the experimental group they represented the familiar. Both the fact that 53% of the control $S$s reached criterion in the first 10 trials as opposed to only 18% of the experimental group, and the greater ease of learning the cross-hatched ground cue are subject to the same

interpretation. It might also be noted that the greater ease in learning the Stage III task where the positive Stage I cue was present as a negative instance fits well with this interpretation. This familiar cue constituted a clearly defined negative instance 50% of the time and thereby served to orient $S$ to the other card which was now a positive instance. Had the primary cue been randomly associated with either positive or negative instance in Stage III it would not have had the effect of increasing novelty and, therefore, discriminability of the positive instance.

Two further task variables are probably crucial to the evaluation of the effects of the acquisition of initially irrelevant information. One is that positive transfer of previously irrelevant cues may occur (Babb, 1956; Bruner et al., 1955; Lawrence, 1949, 1950) only where the irrelevant stimuli are not competing in the same sensory field, e.g., a main visual task with irrelevant auditory cues. Different sensory channels would seem to reduce cue interference and the necessity of "adapting out" (Restle, 1955) cues. The second factor is that where the irrelevant cues are randomly present throughout a Stage I task, and then systematically related to the positive cue in Stage II only for the experimental group, then the novelty advantage of these cues for the control group in Stage III should be nullified.

## SUMMARY

A three-stage concept-learning task was used to investigate the effects on subsequent cue utilization of, prior exposure to relevant cues, differences in level of motivation ($MA$ scale), and the shift of a previously positive cue to the status of a negative instance 50% of the time. The experiment was in the form of a $2 \times 2 \times 2$ factorial design with 10 $S$s per cell.

The results showed no differences on any aspect of the concept-learning task as a function of level of motivation. The introduction of additional (but irrelevant) cues, after an initial discrimination has been learned, function so as to inhibit rather than facilitate the utilization of these cues when they later become the basis for correct

discrimination. A previously positive cue retained in a later learning task, where its reinforcement value is negative, facilitates the learning of a new discrimination. Results were discussed as suggesting an interpretation in terms of stimulus novelty.

## References

Babb, H. Proportional reinforcement of irrelevant stimuli and transfer value. *J. comp. physiol. Psychol.*, 1956, 49, 586–587.

Berlyne, D. E. The influence of complexity and change in visual figures on orienting responses. *J. exp. Psychol.*, 1958, 55, 289–296.

Blum, R. A., & Blum, J. S. Factual issues in the continuity controversy. *Psychol. Rev.*, 1949, 56, 33–50.

Bruner, J. S., Matter, J., & Papenek, N. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, 62, 1–10.

Cameron, N., & Magaret, A. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

Dember, W. N. *Psychology of perception.* New York: Holt, 1960.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, 66, 183–201.

Herring, F. H. An evaluation of published short forms of the Wechsler-Bellevue scale. *J. consult. Psychol.*, 1952, 16, 119–123.

Jeeves, A. A., & North, A. J. Irrelevant or partially correlated stimuli in discrimination learning. *J. exp. Psychol.*, 1956, 52, 90–94.

Kausler, D. H., & Trapp, E. P. Motivation and cue utilization in intentional and incidental learning. *Psychol. Rev.*, 1960, 67, 373–379.

Lashley, K. S. An examination of the "continuity" theory as applied to discrimination. *J. gen. Psychol.*, 1942, 26, 241–265.

Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, 39, 770–784.

Lawrence, D. H. Acquired distinctiveness of Cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, 40, 175–188.

Luria, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957. Pp. 115–129.

Peterson, L. R., & Peterson, M. J. The role of context stimuli in verbal learning. *J. exp. Psychol.*, 1957, 53, 102–105.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, 62, 11–19.

Sokolov, E. N. Higher nervous energy and the problem of perception. In, *Communications to the XIV International Congress of Psychology*, USSR, 1954.

Sullivan, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

Taylor, J. A. The Taylor manifest anxiety scale and intelligence. *J. abnorm. soc. Psychol.*, 1955, 51, 347–354.

Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, 55, 189–208.

Weiss, W., & Margolius, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, 48, 318–322.

# EFFECT OF STIMULUS CONDITION AND REACTION TIME INFORMATION ON SPATIAL STIMULUS GENERALIZATION [1]

CHARLES Y. NAKAMURA AND JAQUES W. KASWAN

*University of California, Los Angeles*

Instructed voluntary response (VR) procedures for studying spatial generalization (Brown, Bilodeau, & Baron, 1951) have recently received much attention. But as noted by Sherman and Knopf (1960) little has been done to define the variables that determine error frequency, the measure of stimulus generalization (SG) in this procedure. Although there have been some attempts to generalize formulations derived from classical conditioning situations to the VR situation, there is little agreement between results obtained from the two procedures. One proposed reason for this is that conditioning studies of SG were in large part concerned with primary stimulus generalization, whereas the VR studies must give greater consideration to mediated generalization variables (Gibson, 1959; Mednick & Freedman, 1960).

The concept of stimulus categorization appears to be a useful description of mediational processes involved in VR SG. Used this way categorization involves the assumption that if stimuli to which Ss are instructed to respond can be discriminated and conceptualized as belonging to a common class, this will

facilitate categorization of these stimuli. Where categorization is so facilitated, generalization to other stimuli in the same continuum will be less likely to occur than when categorization is more difficult. The reasoning is similar to Wallach's (1958) discussion of categorization as a determinant in SG-like behavior, where he implied that ease of discrimination of stimuli is directly related to categorizability and inversely related to SG. A recent study by Evans (1961) provides direct evidence that spatial stimulus generalization is a function of whether Ss can perceive the relative spatial location of the training and test stimuli.

A major goal of the present study is to attempt the identification of some stimulus and response variables necessary to specify the role of categorization in VR spatial SG. A brief description of the experimental situation will assist in making the following discussion of hypotheses more explicit. The stimuli are a horizontal row of 11 lights at eye level on a circular panel presented under three different stimulus conditions. In Cond. 5-6-7, Ss are instructed to press a key only when Lights 5, 6, or 7 (positive lights) come on. In Cond. 3-6-9, they are to respond only to Lights 3, 6, or 9. In the third condition, essentially a control group, only the center Light 6 is positive.

Because of the spatial contiguity of Lights 5-6-7, Ss are expected to identify each as belonging to a group of three, thus facilitating categorization of these lights as positive. No such grouping is likely to be discerned in Lights 3-6-9 so that categorization of these lights should be more difficult. Johnsgard (1957), in a study of stimulus-back-

ground contrast, presented a similar explanation of SG effects.

This arrangement of training and test stimuli also permits the comparison of error frequencies predicted by the categorization hypothesis and those predicted by Hull's gradient summation hypothesis based on the exponential summation of overlapping gradients of generalized habit strength. Bilodeau, Brown, and Meryman (1956) reported support for the latter hypothesis. Kalish and Guttman (1957), on the other hand, found only partial support for the Hullian hypothesis both in their review of the Bilodeau et al. (1956) results and in their own research. Under the conditions of the present experiment, the gradient summation hypothesis would predict approximately the same number of errors to Lights 4 and 8 in both Cond. 5-6-7 and 3-6-9. (Lights 4 and 8 were selected for specific comparison of errors because in both stimulus conditions they are located ideally in the spatial continuum where there are at each negative light three overlapping gradients generated from the training lights.) If, however, spatial contiguity facilitates categorization, fewer errors to these two lights would be expected in 5-6-7 than in 3-6-9. Moreover, in accord with the categorization hypothesis that the three positive lights in 5-6-7 are conceptualized as belonging to a single class, both the total number of errors and the error gradients in 5-6-7 are expected to be similar to those obtained in Stimulus Condition 6 where only the single light Number 6 is positive. Such a prediction would not be made from the gradient summation hypothesis.

Specific predictions of response latencies also follow from the categorization hypothesis. On the assumption that reaction time (RT) generally increases as a function of stimulus complexity (Flores, 1956; Grebb, 1954) and that categorization is presumably a determinant in accuracy of perception (Bruner, Goodnow, & Austin, 1956, p. 9), it would be expected that accurate response latencies should be shorter in 5-6-7 than in 3-6-9.

The above predictions of errors and latencies generated from the stimulus categorization hypothesis produce, as logical consequents, some specific predictions of error-latency relations. If latencies reflect difficulty of categorization of positive lights, Ss who do not take the required time for accurate categorizing should make more errors than those who take adequate time. Accordingly, if Ss can be induced to reduce response latencies they should make more errors. Since there is some evidence that knowledge of performance may reduce RT (Bilodeau & Bilodeau, 1961, p. 250), giving Ss information about their speed of reaction should reduce latency. Insofar as 5-6-7 latencies are expected to be close to simple RTs, further reductions in RT or increase in errors should be limited. In 3-6-9, however, greater absolute reductions in latencies are possible. Accordingly, greater amount of speed information should be associated with larger error differences and smaller latency differences between the two stimulus conditions, and these should be attributable to the speed information effect in Cond. 3-6-9.

## METHOD

*Apparatus.*—The apparatus was a modification of that used by Brown et al. (1951). The main component was a 180° semicircular panel on which was mounted a horizontal row of 11 lights separated by intervals of 4° of visual angle, at eye level. The essential modification was the use of 4° of visual angle with the lights at a distance of 30 in. from S's eyes instead of the 8° intervals at a distance of 60 in. employed by Brown et al. The lights, identified by numbering 1 through 11 from left to right, were standard 6.3-v., .15-amp. miniature bayonet indicator lamps encased in milk-glass covers. A similar orange colored light located directly above the central light, Light 6, served as both fixation point and ready signal. The panel was 3 ft. high, painted flat black, covered over the top with black cloth, and placed in a semidark room to minimize extraneous visual cues. The S responded by pressing a telegraph key placed at his right hand (or left, if left handed). Trial intervals were regulated by an electric

motor. The $E$, seated behind the panel and not visible to $S$, selected the stimulus light, randomly varied the foreperiod between 2, 3, and 4 sec. on a Hunter electric interval timer, and read the RTs to the nearest .01 sec. from a Hunter electric clock.

*Procedure.*—The basic 2 × 3 factorial design involved six sets of $S$s under two stimulus conditions and three RT information groups. The groups under Cond. 5-6-7 were instructed to respond to the lighting of either Lights 5, 6, or 7 and those in Cond. 3-6-9 to either 3, 6, or 9. All the groups were instructed not to respond when any of the other lights were lit.

Within Cond. 5-6-7 and 3-6-9, $S$s in the total information (TI) group were given information of speed of response after every trial and the partial information (PI) group after every fifth trial. The $S$s were told their performance was "good" (short latency), "fair," or "poor" (long latency) in equal proportion and $E$ attempted to make this correspond with $S$s' actual variations in latencies. For the no information (NI) group, speed and accuracy were stressed equally and instructions to this effect were repeated after every 25 trials. All $S$s were told when they made an error.

After the sequence of ready signal, stimulus light, and $S$'s response was demonstrated, $S$ was given a series of 20 training trials distributed among the three positive lights in a predetermined random sequence. Trial intervals were 12 sec., beginning with the onset of the ready light. Following the training trials, $S$ was told that now any of the lights might go on but he was to respond only to the positive lights (5, 6, 7, or 3, 6, 9, depending on the stimulus condition) and not to any other lights. The test series involved 5 presentations of each of the negative lights, along with 40 booster trials of each of the three positive lights distributed in a predetermined random order.

In addition to Cond. 5-6-7 and 3-6-9 in the basic design, a third stimulus condition (Cond. 6) was included where $S$s were instructed to respond only to Light 6. This group was given total RT information similar to the other TI groups. The number of booster trials given Light 6 during the test series was equal to the sum of booster trials to the three positive lights in each of the other two stimulus conditions.

*Subjects.*—The $S$s were 130 students, both men and women, from a course in introductory psychology. Within each of Cond. 5-6-7 and 3-6-9, 20 $S$s were randomly assigned to each PI and TI group and 15 to each NI speed

information group. There were 20 $S$s in Cond. 6.

## RESULTS

*Errors.*—The mean percentages of errors to each of the negative lights are plotted for the different speed information conditions in Fig. 1 for 5-6-7 and Fig. 2 for 3-6-9. The analysis of variance of total errors over all information groups showed that the stimulus condition effect was significant ($F = 38.19$, $df = 1/104$, $P < .001$) with fewer total errors in 5-6-7 than in 3-6-9. The stimulus effect was also significant for errors to Lights 4 and 8 ($F = 14.91$, $df = 1/104$, $P < .001$), with the fewer errors in 5-6-7. This is in contrast to the nearly identical averages of the summation hypothesis predictions to these two lights of 54% and 52% errors
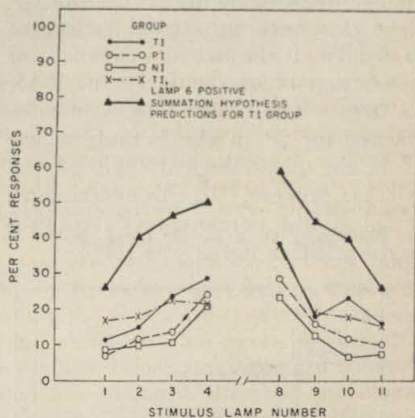


FIG. 1. Mean percentage of errors in Cond. 5-6-7 to each negative (test) lamp, and percentages predicted by the summation hypothesis for the TI group. (The latter were obtained by erecting around each of Lights 5, 6, and 7 the gradient obtained when only Light 6 was positive, and exponentially summating overlapping points. The two wings of the single Light 6, also shown, are displaced one spatial unit away from center so that there is correspondence of percentage of errors to each test light, for all gradients, on the dimension of distance from the nearest positive light.)

expected for 5-6-7 and 3-6-9, respectively. (Lights 4 and 8 frequencies were combined in the analysis of variance since a chi square test of the difference between errors to these lights was not significant—$\chi^2 = .90$, $df = 2$, $P > .10$.) These results are consistent with the stimulus grouping hypothesis that the spatial contiguity of positive lights facilitated categorization and led to the reduction of errors. Errors are clearly not attributable to a failure to discriminate lights as positive since $Ss$ failed to respond appropriately to positive lights in only three trials. The hypothetical summation gradients in Fig. 1 and 2 were plotted by erecting around each of Lights 5, 6, and 7 and around Lights 3, 6, and 9 the gradient obtained when only Light 6 was positive. The values of the overlapping points, at each negative light, were summated according to Hull's formula for exponential summation of gradients (Hull, 1943, p. 200). The value 100 was substituted for $M$ in the formula where $M$ is the physiological limit of the learning process. In this experiment, the maximum percentage of responses to the positive lights was 100%.
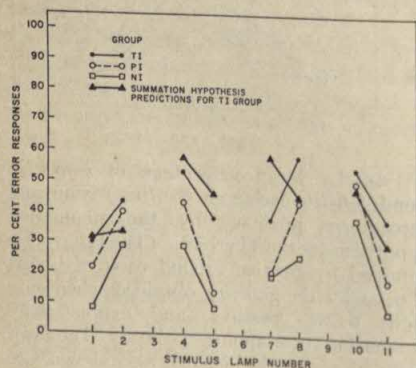


Fig. 2. Mean percentage of errors in Cond. 3-6-9 to each negative (test) lamp, and percentages predicted by the summation hypothesis for the TI group. (The latter were computed as described for Fig. 1.)
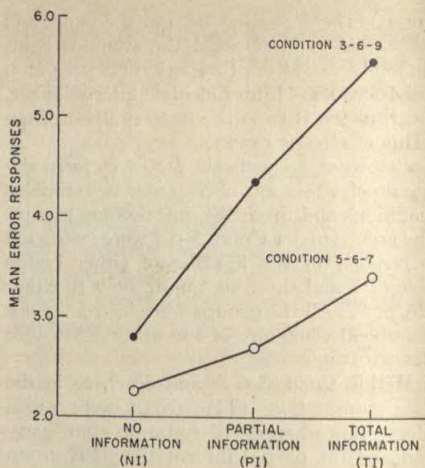


FIG. 3. Mean number of errors to Lights 4 and 8 in two stimulus conditions and three speed information groups.

Figure 1 shows that the error gradient of the TI group in Cond. 5-6-7 closely approximated the gradient obtained in Cond. 6, also under TI. The mean difference between these sets of points was not significant ($F = 1.57$, $df = 1/38$), indicating that the number of stimuli to be categorized did not affect generalization. Moreover, the large discrepancy between summation hypothesis predictions and obtained error frequencies for the TI group in Cond. 5-6-7 indicates that summation does not account for the results obtained by this VR procedure. The mean difference between these two sets of points was significant ($F = 23.81$, $df = 1/8$, $P < .01$).[2] However, Fig. 2 suggests

[2] In order to obtain scores for a variance estimate of the hypothesized gradient, necessary to the analyses of differences between distributions, the summation procedure was carried out separately for Cond. 5-6-7 and 3-6-9 by using the error scores, to the relevant lights, made by each $S$ in Cond. 6. A statistical problem posed because several $Ss$ in Cond. 6 had zero error scores could not be resolved by a transformation since the subsequent scores were to be exponentially summated. This was overcome by using

that in Cond. 3-6-9 the results obtained for the TI group approximated fairly closely the predictions of the summation hypothesis. The nonsignificant $F$ of 2.39 ($df = 1/8$) for the difference between these curves was consistent with this observation.

The curves in Fig. 3 plot mean errors to Lights 4 and 8 as a function of speed information for Cond. 5-6-7 and 3-6-9 and show that greater amounts of speed information were associated with higher occurrence of errors ($F = 7.39$, $df = 2/104$, $P < .01$). The information effect was also found to be significant for total errors to all negative lights ($F = 12.15$, $df = 2/104$, $P < .001$). A further $t$ test analysis of errors to Lights 4 and 8 at each information level showed that the stimulus condition effect occurred only where infor-

the mean scores of $S$s randomly combined into groups of 4, at the cost of reduction in $df$.

mation was given ($t = 2.56$, $df = 38$, $P < .05$, for the difference between PI groups and $t = 3.24$, $df = 38$, $P < .01$, for the TI groups). The corresponding difference between the NI groups was not significant.

*Latencies.*—All the latency computations were based on the reciprocals of the latency scores (Edwards, 1960, p. 131) and the weight of the statistical analyses was almost wholly on the latency scores to positive lights. The latter was necessary since the latencies for errors were not suited to the usual statistical tests of differences between means since mean latencies for $S$s were computed from unequal numbers of scores because individual $S$s made different numbers of errors to different lights.

Figure 4 shows the marked effect of stimulus condition on latencies to positive lights at all levels of RT informa-
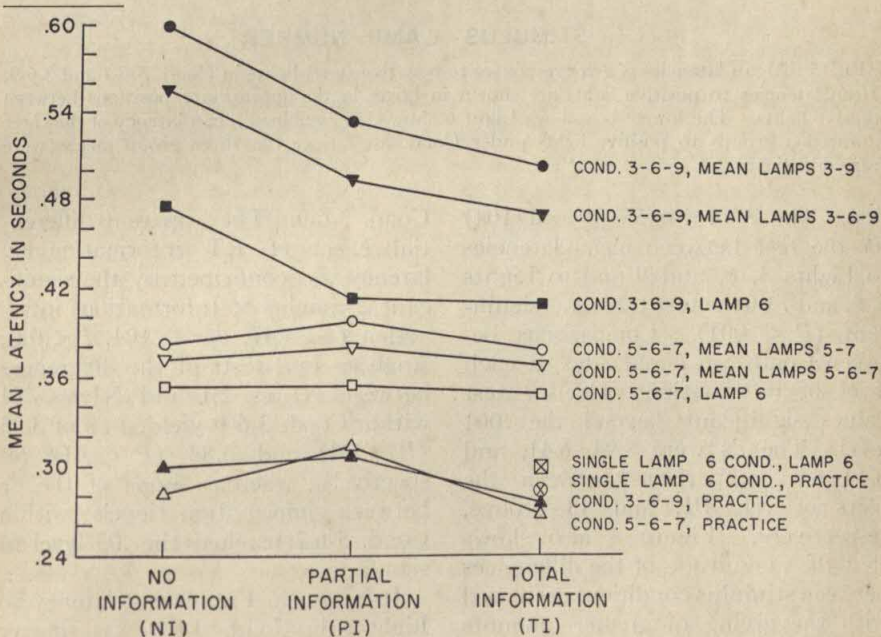


FIG. 4. Mean latencies to positive lights under different stimulus conditions within different speed information groups.
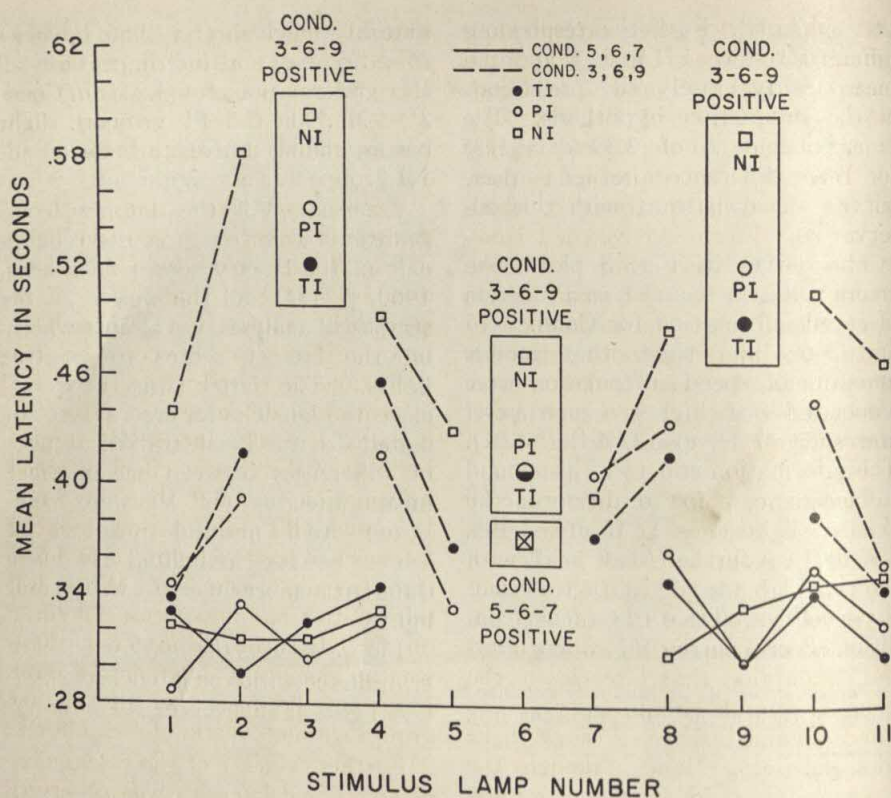
FIG. 5. Mean latencies of error responses to negative (test) lights in Cond. 5-6-7 and 3-6-9. (Mean latencies to positive lights are shown in boxes in the appropriate positions between negative lights. The lowest box above Light 6 shows the combined mean latency of the three information groups to positive lights under Cond. 5-6-7 since the three group means were nearly identical.)

tion. The $F$ of 148.55 ($df = 1/104$) for the test between mean latencies to Lights 3, 6, and 9 and to Lights 5, 6, and 7 of the test trials was significant ($P < .001$). Comparisons between Cond. 5-6-7 and 3-6-9 at each level of information yielded $t$ test values significant beyond the .001 level. The $t$'s were 8.94, 6.41, and 6.49 for comparisons between the pairs of NI, PI, and TI groups, respectively. Figure 4 also shows that the magnitude of the differences between stimulus conditions decreased with the giving of greater amounts of speed information and indicates that the locus of this effect was in

Cond. 3-6-9. The apparent differential effect of RT information on latency was confirmed by the significant Stimulus $\times$ Information interaction ($F = 4.47$, $df = 2/104$, $P < .02$). Analyses by $t$ tests of the differences between NI vs. SI and NI vs. TI within Cond. 3-6-9 yielded $t$'s of 3.05 ($P < .01$) and 3.84 ($P < .01$), respectively; whereas, none of the $t$'s between information levels within Cond. 5-6-7 reached the .05 level of significance.

It is seen in Fig. 4 that latency to Light 6 in Cond. 3-6-9 was shorter than the mean latency to Lights 3 and 9 ($F = 46.84$, $df = 1/104$, $P < .001$)

and similarly for the corresponding difference between Light 6 and the mean for Lights 5 and 7 in Cond. 5-6-7 ($F = 12.99$, $df = 1/104$, $P < .001$). It is probable that the shorter latency for Light 6 is attributable, in part, to the signal light above it which served as a discriminative cue. However, it is unlikely that this alone could account for the main effects since the absolute mean latency to Light 6 differed among stimulus conditions. It was longest under Cond. 3-6-9, shorter in Cond. 5-6-7, and shortest in Cond. 6, as shown in Fig. 4.

Comparison of the mean of the three mean latencies to positive lights for the TI group in Cond. 5-6-7, with the mean latency to Light 6 for Cond. 6 yielded an $F$ of 18.73 ($df = 1/38$, $P < .001$). This finding contradicted the assumption that latencies to the grouped stimuli should approximate those obtained when a single light was positive. Hence, though the error gradients did not differ for these two stimulus conditions, as noted in Fig. 1, the data suggest that categorizing three contiguous stimuli as a class is in fact more difficult than distinguishing a single positive stimulus in a continuum of stimuli.

Figure 5 plots error response latencies to the negative lights for the three RT information groups within each of Cond. 5-6-7 and 3-6-9. The upper part of Fig. 5 shows that in 3-6-9, latencies for PI and TI groups tend to be shorter than those for the NI group. In contrast, (lower part of Fig. 5) there were no substantial differences in latency gradients or in latency to individual lights as a function of speed information in Cond. 5-6-7, and error latencies were clearly much shorter than those in Cond. 3-6-9 with practically no overlap. Also, error latencies were con-

sistently much shorter than latencies to adjoining positive lights for all three information groups within Cond. 3-6-9 and the differences were slight but in the same direction for Cond. 5-6-7.

*Error-latency interaction.* — Comparison of latencies to positive lights in Fig. 4 with errors for Cond. 5-6-7 and 3-6-9 in Fig. 1 and Fig. 2, respectively, shows that although latency differences between the two conditions decreased with greater RT information, differences in error frequency generally increased, largely because of the greater effect of speed information in 3-6-9 than in 5-6-7. Wherever RT information reduced latencies to positive lights, as in 3-6-9, error frequency increased. Where RT information had no substantial effect on such latencies, as in 5-6-7, there was no significant difference in error frequency between the information groups within that stimulus condition.

Further evidence of the interaction of errors and latencies was observed in that all except 1 of the 15 $Ss$ in the NI group of Cond. 3-6-9 obtained mean latencies to positive lights of .5 sec. or greater, whereas, 19 of the 40 $Ss$ given RT information (PI and TI groups) in Cond. 3-6-9 had mean latencies shorter than .5 sec. Since the NI group obtained significantly fewer errors than the PI and TI groups in 3-6-9, the .5-sec. latency may be conceived as an optimum time for categorization of stimulus lights into positive and negative classes. Accordingly, $Ss$ responding faster than that should make more errors than $Ss$ with latencies of .5 sec. or greater. When this comparison was made within TI and PI groups of 3-6-9, an $F$ of 19.85 ($df = 1/38$, $P < .001$) was obtained. Moreover, comparison of error frequency of $Ss$ above and below the median latency of .58 sec. within

the NI group in Cond. 3-6-9 also yielded a significant result ($t = 2.26$, $df = 28$, $P < .05$) for further indication of the relation of shorter latency with greater errors. Similarly for Cond. 5-6-7, though neither error frequency nor latency had differed as a function of RT information, frequency of errors between $Ss$ above and below the median latency of .4 sec., for the three RT information groups combined, was significantly different ($F = 4.19$, $df = 1/53$, $P < .05$).

A third source of evidence for error-latency interaction was the negative Pearson product-moment $r$'s between frequency and latency of errors. For Cond. 5-6-7, the $r$'s for each group were NI = $-.51$, PI = $-.41$, and TI = $-.50$. Under Cond. 3-6-9, they were NI = $-.48$, PI = $-.81$, and TI = $-.71$. All except the $r$ of $-.41$ were significant at the .05 level or better. As expected, errors and latencies were most substantially correlated when RT information was given in Cond. 3-6-9, where stimulus categorization was presumably most difficult. However, the only significant difference ($P < .05$) between correlations between Cond. 5-6-7 and 3-6-9 within any information level was that between the $-.81$ and $-.41$ for the PI groups.

When percentages of errors in Fig. 1 and 2 are compared with error latencies in Fig. 5, the direction of changes in magnitude from light to light are notably parallel. Though this is particularly true of Cond. 3-6-9, it is also seen in Cond. 5-6-7; where decreases in error are greatest, as from Lights 4 to 3 and from Lights 8 to 9, error latencies show a corresponding decrease in five of the six instances. While latency gradients beyond these points are somewhat irregular, variations are within a narrow range suggesting an essentially horizontal

distribution consistent with the flat error frequency gradients. Thus the results suggest that error frequency and error latency *gradients* are *positively* related.

## DISCUSSION

The assumption that stimulus conditions facilitate the categorization of lights as positive or negative was consistently supported. It seems, however, that the stimulus contiguity hypothesis does not entirely account for this facilitation. The persistent sharp drop in errors and latencies to lights adjoining Light 6 in Cond. 3-6-9 and the shorter latency to Light 6, compared to other positive lights, suggest that the central position of this light and (or) the signal light located above it served as discriminative cues which facilitated categorization of the adjoining lights as negative stimuli. It is possible that the central location of the positive lights and the signal light also played a role in Cond. 5-6-7, although the first negative light is two lights removed from the signal light. Occurrence of such effects had not been expected since Bilodeau et al. (1956), using a similar apparatus, found that error gradients and frequencies to single positive lights located 16° of visual angle from center were virtually identical with those obtained to a central positive light. Though this finding clouds the interpretation of results in terms of a stimulus pattern effect, the influence of the latter is apparent in the fact that latencies to Light 6 itself decrease from Cond. 3-6-9 to 5-6-7 to Cond. 6, suggesting an effect of stimulus arrangement on speed of categorization.

Specification of stimulus arrangement and stimulus position as determinants of SG appear of special interest because this involves the problem of specifying the role of stimulus units in SG. In their review, Mednick and Freedman (1960) indicate that most SG studies follow an empirical formulation which assumes that SG is a decreasing function of increasing physical difference between

test and training stimuli. Hull (1943, p. 198) explicitly recognized the untenability of postulating such arbitrary simple physical stimulus units and proposed the jnd as the stimulus unit for establishing SG gradients. The present results re-emphasize that equal physical distance units from different positive lights do not correspond to equal decrements in SG in the VR paradigm. This was shown most explicitly by the comparison of the error frequencies obtained in Cond. 5-6-7 with the higher frequencies of Cond. 3-6-9, and between the obtained curves and the summation hypothesis predictions in each of these conditions. It was further demonstrated by the fact that error frequencies to lights adjoining Light 6 in Cond. 3-6-9 were much lower than errors to lights surrounding Light 3 and Light 9, although the difference in physical distance was identical. The results indicate the advantage of a stimulus unit or scale that could take into account the specification of the relation among stimuli. The proposal of an ordinal scale of stimulus values by Mednick and Freedman (1960) is an attempt in this direction. However, their approach would predict sharper gradients over the negative lights in Cond. 5-6-7 than obtained in the present study. Since Lights 4 and 8 are separated from the training stimuli by only one ordinal unit, Mednick and Freedman would expect higher percentages of errors on them, relative to the more peripheral lights, because their formulation does not take into account the effect of contiguity of the training lights.

The overall results of this study show a close relation between error and latency. The negative error-latency relation wherein Ss who took less time before responding to training stimuli generally had greater error frequencies indicates that errors were due largely to insufficient time taken (for accurate categorization) before responding. Another aspect of the close relation between categorization and latency is observable in the positive relation between error frequency and error latency which is also found in most other VR SG studies

(Gibson, 1939; Mednick, 1958; Rosenbaum, 1953).

The manifold role of latency in the results of this study make it reasonable to suggest that latency may serve to specify both stimulus and response anchors of the term "categorization" used here to describe the verbal and other mediated processes which are undoubtedly important determinants of SG in any VR paradigm. Stimulus referents of categorization can be specified by noting that the time taken for responses varied with the arrangement or position of lights, as in the different latencies for Cond. 5-6-7 and 3-6-9. Response referents of categorization can be specified by noting that latency was a major determinant of error frequency under given stimulus arrangement conditions, as in the effect of speed information in Cond. 3-6-9. Thus, difficulty or ease of categorization is measurable in terms of latency required to make a correct response, under specified conditions such as, in this study, relative emphasis on speed induced by providing reaction time information.

## SUMMARY

This study examined the effects of stimulus arrangement and experimentally induced differences in response latency on performance in a spatial stimulus generalization situation involving voluntary responses. In Cond. 5-6 7, Ss were instructed to respond to any one of three centrally and contiguously located lights (Lights 5, 6, and 7) of 11 lights in a row, but to inhibit responses to all other lights. In Cond. 3-6-9, the positive lights were noncontiguous (Lights 3, 6, and 9). The results showed that error frequency and latency were smaller in Cond. 5-6-7 than in Cond. 3-6-9. It was also found that giving Ss information on speed of responses reduced latencies and increased errors in Cond. 3-6-9 but had practically no effect in Cond. 5-6-7. These findings were consistent with predictions of stimulus generalization generated from a stimulus categorization hypothesis. Predictions of error frequency by this hypothesis were compared with those made from Hull's gradient summation hypothesis of stimulus generalization. The latter failed to predict any of the obtained points on the gradient for Cond. 5-6-7, but did approximate

most of the points obtained in Cond. 3-6-9, under conditions where $S$s were given maximum speed information. The close and consistent relationship found between error frequency and latency was interpreted as indicating that latency can be used to specify both stimulus and response anchors of the mediational process in spatial VR SG.

## REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, 12, 243–280.

BILODEAU, E. A., BROWN, J. S., & MERYMAN, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, 51, 293–298.

BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, 41, 52–61.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, J. A. *A study of thinking.* New York: Wiley, 1956.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

EVANS, W. O. Two factors affecting stimulus generalization on a spatial dimension. *J. exp. Psychol.*, 1961, 61, 142–149.

FLORES, I. The effect of organization upon complex reaction time. *J. Psychol.*, 1956, 41, 301–313.

GIBSON, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, 24, 237–253.

GIBSON, E. J. A re-examination of generalization. *Psychol. Rev.*, 1959, 66, 340–342.

GREBB, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, 48, 289–297.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JOHNSGARD, K. W. The role of contrast in stimulus intensity dynamism (*V*). *J. exp. Psychol.*, 1957, 53, 173–179.

KALISH, H. I., & GUTTMAN, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, 53, 139–144.

MEDNICK, S. A. Gradients of latency in a generalized voluntary response. *Amer. J. Psychol.*, 1958, 71, 752–755.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, 57, 169–200.

ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, 45, 35–43.

SHERMAN, J., & KNOPF, I. J. Changes in the gradient of stimulus generalization as a function of two procedural variations. *Psychol. Rep.*, 1960, 7, 253–258.

WALLACH, M. A. On psychological similarity. *Psychol. Rev.*, 1958, 65, 103–116.

# STIMULUS GENERALIZATION AS A FUNCTION OF THE FRAME OF REFERENCE [1]

DAVID R. THOMAS AND CHARLES G. JONES

*Kent State University*

In a study reported by Philip (1952) it was shown that the location of the original stimulus within a generalization test series modifies the shape of the obtained gradient. Philip required his Ss to rank cards containing varying proportions of green and blue dots along a greenness-blueness scale. The frequency with which a given rank, say Number 3, was subjectively attributed to the different cards constituted a gradient of generalization for that value. A separate gradient was thus generated for each rank employed.

Philip systematically varied the length of the generalization test series. With the shortest series employed (six values) there was a tendency for judgments to accumulate near the center of the scale, the "central tendency effect" (Hollingworth, 1909). This effect was reflected in asymmetrical gradients around stimulus values which were non-centrally placed in the series of generalization test stimuli.

Because of the unusual nature of Philip's (1952) procedure, the significance of his finding with regard to generalization as studied by other methods may be questioned. The purpose of the present study was to assess the generality of Philip's finding, using a method for obtaining generalization gradients developed by Kalish (1958). The Kalish procedure is more typical of generalization studies in that Ss are first exposed to a single stimulus value and are subsequently tested for their ability to select the original from a randomly

presented series of stimuli. This is in contrast to the Philip procedure in which absolute judgments of predominating color are made without previous exposure to some standard stimulus value. It was reasoned that if the "central tendency effect" were shown to distort measures involving retention as well as absolute judgment, its relevance for studies of generalization would be more convincingly demonstrated.

## METHOD

*Subjects.*—The Ss were 50 undergraduate men taken from introductory psychology courses at Kent State University. All Ss had normal color vision, as determined with the Dvorine (1944) color perception test.

*Apparatus.*—The study employed a Skinner-type key pecking apparatus, modified for use with human Ss. The box was approximately 15 in. long, 11 in. high, and 14 in. wide and was painted flat black. The front wall of the box was made of transparent Plexiglas so that Ss could clearly view the pecking key. The S sat in a chair approximately 2½ ft. from the key, which was a circular plastic disc ⅜ in. in diameter. Illumination was provided by a Cambridge Thermionic Corporation monochromator, Model B, Series 1066, equipped with an Olympus, Model 201250 6-v., 5-amp. light source. The patch of color on the key was approximately 4 mL. in luminance. The only other light in the room was a 7.5-w. "night light" on E's side of a black cloth screen separating S from E.

A telegraph key was used to measure S's responses. The telegraph key was placed on the table next to the Skinner box, within easy reach of S's right hand. It was wired so that its release would illuminate a signal light on E's side of the screen, thus signifying a response. The box was equipped with an electrically operated shutter which interrupted the monochromator beam when E threw a switch.
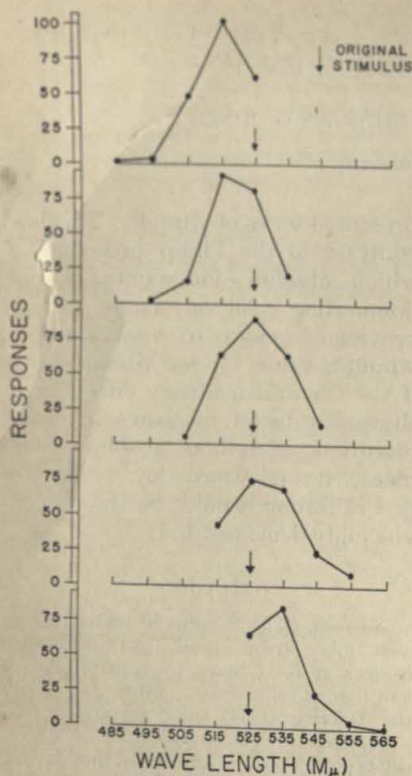
FIG. 1. Generalization gradients of the five experimental groups. (The gradient of Group 1 is at the top of the figure with that of Group 2 directly below it, etc. Note that the value of the original stimulus is the same for all groups.)

*Procedure.*—The Ss were divided unsystematically into five groups of 10 Ss each. Each S received the same instructions. They were as follows:

This is an experiment in color perception. At the beginning of the experiment a specific color will be presented through the small hole in front of you. Try to keep this color in mind because you will be asked to identify it later. After 1 minute this color will be turned off and you will place your finger and press down on the telegraph key in front of you. I will give the signal "ready" and a few seconds later a color will again be presented. You must decide whether this is the original color shown you at the start of the experiment. If it is, *lift* your hand as rapidly as you can

from the key. If it is not, keep pressing on the key.

I will say the word "ready" whenever I am about to present a color and you should be pressing the key at that time. We are going to try some practice trials. Now we are going to run through a series exactly as we would do if this were the real experiment.

At this point the instructions were interrupted and a stimulus of 600 mμ was presented for 1 min. Then S was presented with 610 mμ, 590 mμ, 600 mμ, 620 mμ, and 580 mμ. Each test stimulus was presented for 5 sec. with 5 sec. intervening between presentations. If S appeared to have understood the instructions, they were continued as follows:

Now we are going to begin the experiment. Remember, try to keep the original color in mind and respond as rapidly as you can, lifting your finger only when the original color appears. Do not be disturbed, however, if you should respond to other colors.

The Ss in all five groups were presented with the same original stimulus, 525 mμ (a middle-green), for 60 sec. The groups differed only with regard to the series of stimuli employed in testing for generalization. In Group 1 the test stimuli were 485 mμ through 525 mμ, in 10-mμ steps. For Group 2 the range covered was 495 mμ–535 mμ, for Group 3, 505 mμ–545 mμ, for Group 4, 515 mμ–555 mμ, and for Group 5, 525 mμ–565 mμ. For each S the five test stimuli were randomized within a series and 12 different series were presented. The number of responses made to the different test stimuli constituted a generalization gradient.

## RESULTS AND DISCUSSION

In Fig. 1 are presented the generalization gradients of the five groups of Ss. It should be remembered that all groups were exposed to the same standard stimulus. In spite of this, the five gradients differ strikingly in a manner consistent with the findings of Philip (1952). The tendency to respond to stimuli closer to the center of the test series is so strong that the peak of the generalization gradient tends to be displaced from

the value of the original stimulus. Thus, a change of, for example, 10 m$\mu$ from the original stimulus does not produce a fixed generalization decrement, but may result in no change in response strength or even in an increment depending on the location within the range of test stimuli. It should be noted that only with Group 3, where the test stimuli were symmetrically distributed around 525 m$\mu$, did the peak of the gradient fall clearly at that value!

The difference in gradient shape was shown to be statistically reliable in the following manner: A simple analysis of variance was performed to test for differences in the mean number of responses given to 515 m$\mu$ in Groups 1, 2, 3, and 4. The result was an $F$ of 9.06 ($df = 3/36$, $P < .01$). A parallel analysis was performed for mean responses to 535 m$\mu$ in Groups 2, 3, 4, and 5. The result was an $F$ of 8.41 ($df = 3/36$, $P < .01$).

The purpose of this study was to determine whether the effect of "central tendency" on the generalization gradient is peculiar to the absolute judgment situation employed by Philip. It is safe to conclude that it is not. Indeed, the distortion of the gradient proved even greater in the present experimental situation than under the conditions of Philip's experiment.

These findings may be interpreted with reference to Helson's (1947) theory of adaptation level. It may be argued that the series of test stimuli provides a frame of reference against which the memory trace of the original stimulus is judged. When test stimuli are presented which fall asymmetrically around the original stimulus, a change in the frame of reference may be assumed to result, culminating in a heightened tendency to respond to stimuli nearer to the center of the test series, thereby distorting the resulting generalization gradient.

The significance of the "central tend-

ency effect" for studies of generalization employing conditioning techniques remains an open question. We would guess, however, that the background of test stimuli would exert far less influence in that situation than in the present one. In this study, $S$'s experience with the original stimulus was limited to one 60-sec. exposure. The resulting limited familiarity with the stimulus makes extensive retroactive interference by the test series which follows more likely. In a conditioning situation, however, $S$ has the opportunity through repeated exposure to become much more familiar with the value of the CS. Greater familiarity with the stimulus should reduce the effect of the test situation which is later employed. Some direct evidence on this issue has been reported. Guttman (1956) discussed some pilot work with pigeons and an operant conditioning technique in which the asymmetry of the distribution of test stimuli appeared to have no effect on the resulting generalization gradient. More recent work from the Kent State laboratory has tended to corroborate his finding. Of course the species-related and procedure-related differences between the pigeon and human studies are so great as to make these findings merely suggestive. Research is needed to assess the role of familiarity with the CS in determining the influence of the "central tendency effect" on generalization.

## SUMMARY

Five groups of 10 $S$s each viewed a monochromatic light of 525 m$\mu$ (a middle-green) for 60 sec., and then were presented 12 different random series of wave lengths under instructions to respond only to the original color. The number of responses made to the different test stimuli constituted a gradient of generalization. Group 1 was tested with the series 485–525 m$\mu$, in 10-m$\mu$ steps; Group 2, 495–535 m$\mu$; Group 3, 505–545 m$\mu$; Group 4, 515–555 m$\mu$; and Group 5, 525–565 m$\mu$. Only Group 3, with a central value of 525 m$\mu$, produced a generalization gradient with a definite peak at 525 m$\mu$; in all other cases the peak of responding was displaced toward the center of the series of test stimuli. The extent

of this displacement varied directly with the degree of asymmetry of the test series around the value of the original stimulus. These results support the assumption that the generalization test series serves as a frame of reference against which the memory trace of the original stimulus is judged.

## References

Dvorine, I.  *Dvorine color perception testing charts*.  Baltimore, Md.: Waverly, 1944.

Guttman, N.  The pigeon and the spectrum and other perplexities.  *Psychol. Rep.*, 1956, 2, 449–460.

Helson, H.  Adaptation level as a frame of reference for prediction of psychophysical data.  *Amer. J. Psychol.*, 1947, 60, 1–29.

Hollingworth, H. L.  The inaccuracy of movement.  *Arch. Psychol.*, *N. Y.*, 1909, 2 (Whole No. 13).

Kalish, H. I.  The relationship between discriminability and generalization:  A re-evaluation.  *J. exp. Psychol.*, 1958, 55, 637–644.

Philip, B. R.  Effect of length of series upon generalization and central tendency in the discrimination of a series of stimuli.  *Canad. J. Psychol.*, 1952, 6, 173–178.

# EFFECT OF REWARD MAGNITUDE, PERCENTAGE OF REIN-FORCEMENT, AND TRAINING METHOD ON ACQUISITION AND REVERSAL IN A T MAZE [1]

WINFRED F. HILL

*Northwestern University*

JOHN W. COTTON

*University of California, Santa Barbara*

AND KEITH N. CLAYTON

*Vanderbilt University*

The effect of reward magnitude on extinction is an unresolved question. Zeaman (1949) and Metzger, Cotton, and Lewis (1957) have found faster running early in extinction for animals receiving the larger acquisition reward, but convergence of the curves as extinction progressed. This is consistent with the assumption (Hull, 1952; Spence, 1956) that the incentive motivation factor ($K$) adjusts rapidly but not immediately to a change in reward magnitude. On the other hand, Hulse (1958) and Armus (1959) have found faster running over a number of extinction trials for animals receiving the smaller reward. This might be predicted from the depression effect found in some studies of magnitude change during acquisition (Spence, 1956), since no reward is a greater reduction from large reward than from small, but the generality of this effect is open to question. All of the above findings were with 100% reinforcement in acquisition. Hulse's study further complicated the picture by showing an interaction between magnitude and percentage of reward, with large reward producing greater

resistance to extinction than small if rewards were intermittent during acquisition.

The present study translates the question of the effect of reward magnitude on extinction into a special case of extinction—discrimination reversal. In particular, it tests whether the interaction between magnitude and percentage of reward found by Hulse in straight-alley extinction will also appear in reversal of a T maze habit.

While making this test, it is possible in the same experiment to study the effect of reward magnitude on acquisition. A number of prior experiments suggest that larger rewards lead to faster learning of a discrimination if the noncorrection method is used but not if the correction method is used, either in a single-choice situation or in a multiple-unit maze. McKelvey (1956) investigated this apparent interaction of magnitude and method directly and found no effect of magnitude on correct choices with either method. However, he manipulated magnitude by varying the time that rats were permitted to eat rather than the amount of food available. Since rats learn to speed up their eating over a period of deprivation, it is possible that his "large reward" and "small reward" groups may have

81

been eating nearly equal amounts at the end of acquisition. This possibility is strengthened by the fact that an initial difference in speed between the two groups disappeared by the end of acquisition. The acquisition phase of the present experiment incorporates the same basic design as McKelvey's, but with reward magnitude defined by quantity rather than by eating time.

## METHOD

*Subjects.*—The *S*s were 96 female albino rats of the Sprague-Dawley strain purchased from Holtzman Rat Company, Madison, Wisconsin. Eight additional *S*s were discarded because of failure to run, refusal to eat the reward pellets, or *E*'s error. The *S*s were between 72 and 86 days old at the beginning of training.

*Apparatus.*—The *S*s were trained in an enclosed, single-unit T maze. This consisted of a start box 10 in. long and 6 in. wide, a stem 48 in. long and 4 in. wide, arms 7 in. long and 4 in. wide, and goal boxes 13 in. long and 6 in. wide. An alcove for food opened off each goal box in such a way that the reward was not visible from the entrance to the goal box. The entire maze was $8\frac{1}{4}$ in. high, with sides and floor of plywood and lids of Plexiglas. The inside was painted flat black. A guillotine door separated the start box from the stem, and sliding doors separated the stem from the arms. A Standard Electric timer was started by *S*'s weight on a treadle just beyond the start box and stopped by *S*'s weight on either of two treadles just beyond the guillotine doors.

*Design.*—All *S*s received 48 acquisition trials followed by 24 reversal trials. The design was a factorial combination of four dichotomous variables: (*a*) percentage reinforcement of the correct side during acquisition (100% vs. 50% random), (*b*) magnitude of reward during acquisition (four pellets vs. one), (*c*) acquisition training method (correction vs. forced-trial noncorrection), and (*d*) magnitude of reward during reversal. Thus there were 8 treatment cells of 12 *S*s in acquisition and 16 cells of 6 *S*s in reversal. Half of the *S*s in each cell were initially trained to the right side, half to the left.

It was desirable to hold both number of acquisition trials and proportional number

of acquisition reinforcements constant for both magnitudes and both probabilities of reward. For noncorrection *S*s, this required the use of forced trials. Hence, noncorrection *S*s received a combination of 16 free and 32 forced trials in acquisition, so arranged that half of their acquisition responses were to each side. A given ordinal-numbered trial was either free or forced for all noncorrection *S*s, but each *S* in a given cell had a different sequence of forced right and left turns. All trials by the correction method were free. Reversal training for all *S*s was by the noncorrection method with all trials free and with 100% reinforcement.

The above procedure had the effect that a correction *S* received twice as many reinforcements in acquisition as a noncorrection *S* in the corresponding group. However, since method was of interest in connection with acquisition rather than with reversal, this difference was not crucial. Experiments on the role of incorrect responses in discrimination learning suggest that this procedure should make correction and noncorrection groups more nearly comparable than if number of reinforcements were held constant and number of incorrect responses allowed to vary.

*Procedure.*—Seven days before the beginning of training, *S*s were placed on a feeding schedule of 10 gm. of ground Purina chow a day, which continued throughout the experiment. On 6 of these 7 preliminary days, each *S* received 3 min. of handling and was allowed to eat four of the reward pellets. Water was always present in the living cages and the carrying cages throughout the experiment.

In both acquisition and reversal, *S*s were given six trials a day. Thus there were 8 days of acquisition and 4 of reversal, successive except for 1 nontraining day between Days 6 and 7 of acquisition. The reward pellets were 45-mg. Noyes pellets. At the start of each trial *S* was placed in the start box. When *S* was oriented toward the door, *E* opened it, permitting *S* to enter the stem. In the noncorrection procedure, after *S* entered either arm the sliding door was closed and *S* was confined for approximately 15 sec. (If it was a forced trial, the door to the other arm was already closed.) In the correction procedure, if *S* entered the incorrect arm the door remained open and *S* was permitted to retrace, but when *S* entered the correct arm the door was closed and *S* was confined for 15 sec. In both conditions *S* remained in a carrying cage for about 30 sec. between trials. The *S*s received their daily ration about $\frac{1}{2}$ hr. after training.

## RESULTS

*Acquisition.*—The number of correct choices was tabulated for each *S* in acquisition. Since only the free trials were relevant for the noncorrection *S*s, only the corresponding ordinal-numbered trials were counted for the correction *S*s. Thus each acquisition score represented the number of correct turns out of a possible 16. Three cases of failure to run (out of 1536 trials) were counted as errors.

The effects of the three acquisition variables are shown in the three pairs of curves in Fig. 1, each curve based on 48 rats. A triple-classification analysis of variance, with $df = 1/88$ for each $F$ ratio, indicated that all three main effects were

TABLE 1

NUMBER OF CORRECT CHOICES IN REVERSAL

| %<br>Reinf. | Acquisition<br>Reward | Method | Reversal<br>Reward | Mean | SD |
|---|---|---|---|---|---|
| 100 | 1 | C | 1 | 14.83 | 6.58 |
| 100 | 1 | C | 4 | 19.67 | 1.36 |
| 100 | 1 | N | 1 | 10.17 | 1.94 |
| 100 | 1 | N | 4 | 14.00 | 1.90 |
| 100 | 4 | C | 1 | 17.50 | 2.35 |
| 100 | 4 | C | 4 | 17.83 | 1.75 |
| 100 | 4 | N | 1 | 8.50 | 1.87 |
| 100 | 4 | N | 4 | 14.33 | 3.26 |
| 50 | 1 | C | 1 | 17.17 | 4.02 |
| 50 | 1 | C | 4 | 18.17 | 2.40 |
| 50 | 1 | N | 1 | 12.17 | 3.97 |
| 50 | 1 | N | 4 | 18.50 | 2.07 |
| 50 | 4 | C | 1 | 7.67 | 4.08 |
| 50 | 4 | C | 4 | 10.00 | 5.29 |
| 50 | 4 | N | 1 | 7.83 | 5.34 |
| 50 | 4 | N | 4 | 10.17 | 5.23 |

Note.—$N = 6$ each group.

significant. Acquisition was faster for four pellets than for one ($F = 6.71$, $P = .05$), for 100% reinforcement than for 50% ($F = 16.86$, $P = .001$), and for forced noncorrection method than for correction ($F = 7.38$, $P = .01$). None of the interactions was significant, all four $F$s being less than unity. The Hartley test did not indicate any heterogeneity of variance ($F_{max}$ of 3.64). Although there was a marked negative skew in the distribution of scores, it is unlikely (Boneau, 1960) that this had any substantial effect on the $F$s.

*Reversal.*—The number of correct choices on the 24 reversal trials was tabulated for each *S*. Table 1 presents the means and *SD*s of these scores for the 16 groups. These distributions were more nearly normal than those for acquisition, and the apparent heterogeneity of variance was not significant by Bartlett's test ($B = 19.76$, $df = 15$).

Table 2 summarizes the quadruple-classification analysis of variance of these data. Although all four main effects were significant when tested



FIG. 1. Acquisition for all *S*s, classified according to the three acquisition variables.

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT CHOICES IN REVERSAL

| Source | df | MS | F |
|---|---|---|---|
| Percentage reinforcement (%) | 1 | 84.376 | 6.20* |
| Training method (M) | 1 | 273.376 | 20.07*** |
| Acquisition reward (KA) | 1 | 360.376 | 26.46*** |
| Reversal reward (KR) | 1 | 266.667 | 19.58*** |
| % × M | 1 | 126.040 | 9.70** |
| % × KA | 1 | 330.040 | 24.23*** |
| % × KR | 1 | 2.667 | — |
| M × KA | 1 | 3.374 | — |
| M × KR | 1 | 37.499 | 2.75 |
| KA × KR | 1 | 10.667 | — |
| % × M × KA | 1 | 18.377 | 1.35 |
| % × M × KR | 1 | .167 | — |
| % × KA × KR | 1 | 0 | — |
| M × KA × KR | 1 | .667 | — |
| % × M × KA × KR | 1 | 53.998 | 3.96 |
| Within cells[a] | 80 | 13.620 | |

[a] Used as error term for all F tests.
* P < .05.
** P < .01.
*** P < .001.

against the within-cells variance, the three that involved acquisition variables were overshadowed by interactions. Figure 2 shows the course of reversal learning classified in three different ways so as to clarify these interactions. The effect of acquisition variables on reversal may be summarized by saying that 100% correction $S$s reversed fastest and 50% large-reward $S$s reversed most slowly. Reversal, like acquisition, was faster for the larger current reward.

The number of errors prior to the first correct response in reversal was also tabulated and subjected to an analysis of variance on the three acquisition variables. This measure gave substantially the same results as the total correct reversal responses. The main effects of percentage and method fell short of significance, but the effect of acquisition magnitude remained significant at the .001 level. The interactions of percentage with reward size and percentage with method manifested the same patterns as before and were both significant at the .01 level.

*Speeds.*—Time scores were converted to reciprocals and analyses of variance computed. In a triple-classification analysis for the last four trials of acquisition, with $1/88$ $df$ for each $F$, the correction method gave faster running at the .001 level ($F = 131.43$) and larger reward gave faster running for correction-method $S$s only ($F$ for magnitude-method interaction = 4.70, $P = .05$). No other $F$ was significant. In a quadruple-classification analysis for all 24 reversal trials, with $1/80$ $df$ for each $F$, $S$s receiving the larger current reward ran faster ($F = 10.16$, $P = .01$) and those trained with the correction method ran faster ($F = 8.90$, $P = .01$). No other $F$ was significant.
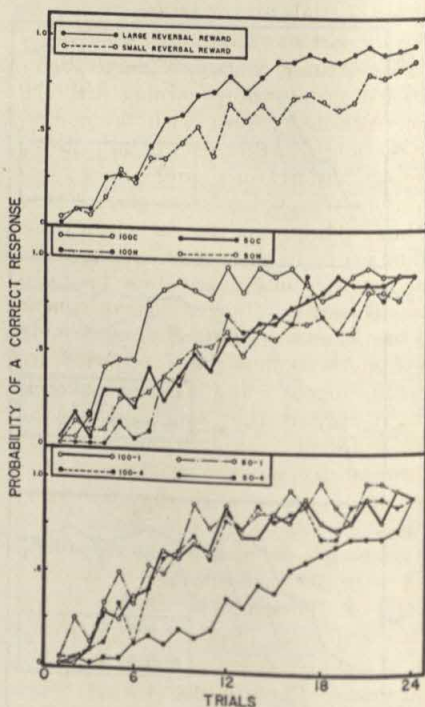


FIG. 2. Reversal learning for all $S$s, classified in three different ways according to reversal variable (top) and acquisition variables (middle and bottom).

## Discussion

*Acquisition.*—The superiority of the forced-trial, noncorrection method, in spite of the smaller number of reinforcements that it provided, argues for the importance in learning of making nonrewarded incorrect responses. Not only did the correction Ss have less experience with the incorrect side, they also were able to obtain reward, although delayed, even when they made the wrong choice. The fact that the noncorrection Ss were forced half the time to the wrong side makes it difficult to compare these results with those of other studies comparing the correction and noncorrection methods. However, in emphasizing the importance of incorrect responses, this study agrees with others that have found better learning of a discrimination with a combination of rewarded and nonrewarded trials than with the same number of trials all rewarded.

No support was found for the hypothesis that greater reward leads to faster spatial discrimination learning with the noncorrection but not with the correction method. Larger reward gave faster learning throughout, and there was no interaction between magnitude and method. The possibility that the present finding is due to the use of forcing in the noncorrection method cannot be ruled out. However, since this experiment was begun, another finding disconfirming the hypothesis has been reported by Lawson, Cross, and Tambe (1959). Thus it appears that some other factor must be found to explain the discrepant prior findings on the relation of discrimination learning to reward magnitude.

The more rapid acquisition with 100% than with 50% reinforcement is comparable to the common, though by no means universal, finding with acquisition of a simple running response.

*Reversal.*—The finding that a large reward 50% of the time produced greater resistance to reversal than any of the other three combinations of magnitude and probability is similar to Hulse's (1958) finding on resistance to extinction

in a straight alley. However, the lack of any difference among the other three groups differs from Hulse's findings. The effect found here is consistent both with a partial reinforcement extinction effect, but only for the large reward condition, and with a greater persistence of more strongly rewarded responses, but only for the 50% condition, and hence is not completely consistent with either of these principles.

Training method had no appreciable effect on reversal when total number of reinforcements was held constant by comparing 50% correction with 100% noncorrection Ss. The interaction that was found between method and percentage may reflect in part a tendency to prefer the less often experienced side (Denny, 1957). Since the 100% correction Ss made the largest number of correct turns in acquisition (the noncorrection Ss being forced equally often to the two sides), this tendency should be greatest for them and should lead to more rapid extinction of the old response and acquisition of the new response, as was found. The partial reinforcement effect might then be invoked to explain why the 50% correction Ss were so markedly slower in reversing than the 100% correction Ss. According to this combination of factors, however, the 50% noncorrection group should have reversed most slowly, which was not the case. Thus this interaction also remains at least partly unexplained.

The faster reversal with larger reversal reward agrees with the finding of faster acquisition with larger reward. Since the reversal training method of free, noncorrection trials was different from either of the acquisition methods, the superiority of the larger reward in this condition increases the generality of the finding that larger reward leads to faster discrimination learning. There was no evidence of generalization decrement resulting from the change in reward magnitude between acquisition and extinction. The lack of interaction between reversal reward magnitude and any of the acquisition variables argues for the independence of acquisition and

reversal variables in their effect on reversal.

*Speeds.*—Since the recorded running speeds include time in the stem plus time in the choice area, the speed data are difficult to interpret. The most striking finding is the superiority of the correction method. McKelvey (1956) also found greater speed and less accuracy with the correction method, and in his study this might be attributed to the greater consistency of reward with correction method. This explanation seems inadequate in the present study, however, since 100% reinforcement did not give significantly faster running than 50%. It is also striking that the superiority of the correction group continues through reversal, even though all *S*s were on noncorrection method during reversal. Since the noncorrection *S*s received primarily forced trials in acquisition, it seems likely that forcing to the incorrect side, rather than noncorrection as such, was the crucial factor. However, why the difference should remain after all animals were changed to free trials is problematic.

In general the effect of magnitude is greater on choices than on speeds. In particular, noncorrection *S*s made more correct choices in acquisition, but did not run faster, for large than for small reward. This argues against Pubols' (1961) suggestion that the effect of magnitude on choices is mediated by the effect on speeds.

## Summary

Acquisition and reversal in a T maze were studied for 96 rats as a function of four variables combined factorially: (a) 100% vs. 50% random reinforcement in acquisition, (b) one vs. four reward pellets in acquisition, (c) correction vs. forced-trial noncorrection method in acquisition, and (d) one vs. four reward pellets in reversal. All *S*s received 100% reinforcement and free-trial noncorrection method in reversal.

Acquisition of the correct response was faster for large reward, 100% reinforcement, and forced noncorrection method, with no interactions. Reversal was faster for large reversal reward, faster after 100% reinforcement with correction method than any other combination of percentage with method, and slower after 50% reinforcement with large acquisition reward than any other combination of percentage with acquisition magnitude. Acquisition by the correction method gave faster running during both acquisition and reversal. Running was faster with larger current reward both during acquisition by correction method and during reversal.

## References

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, 58, 61–63.

Boneau, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, 57, 49–64.

Denny, M. R. Learning through stimulus satiation. *J. exp. Psychol.*, 1957, 54, 62–64.

Hull, C. L. A behavior system. New Haven: Yale Univer. Press, 1952.

Hulse, S. H., Jr. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, 56, 48–57.

Lawson, R., Cross, H. A., & Tambe, J. T. Effects of large and small rewards on maze performance after different prior experiences with reward amounts. *J. comp. physiol. Psychol.*, 1959, 52, 717–720.

McKelvey, R. K. The relationship between training methods and reward variables in brightness discrimination learning. *J. comp. physiol. Psychol.*, 1956, 49, 485–491.

Metzger, R., Cotton, J. W., & Lewis, D. J. Effect of reinforcement magnitude and of order of presentation of different magnitudes on runway behavior. *J. comp. physiol. Psychol.*, 1957, 50, 184–188.

Pubols, B. H. The acquisition and reversal of a position habit as a function of incentive magnitude. *J. comp. physiol. Psychol.*, 1961, 54, 94–97.

Spence, K. W. Behavior theory and conditioning. New Haven: Yale Univer. Press, 1956.

Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, 39, 466–483.

# PAIRED-ASSOCIATE LEARNING UNDER SIMULTANEOUS REPETITION AND NONREPETITION CONDITIONS [1]

WILLIAM F. BATTIG

*University of Virginia*

Several recent investigators (e.g., Clark, Lansford, & Dallenbach, 1960; Estes, 1960; Estes, Hopkins, & Crothers, 1960; Rock, 1957; Rock & Heimer, 1959) have obtained evidence leading them to conclude that associative learning is an all-or-none rather than a gradual incremental process, and that repetition serves only to provide additional opportunities for such all-or-none associations to be learned. However, none of these studies have been sufficiently free of defects in experimental design and procedure to preclude alternative interpretations, discussed at length elsewhere (Postman, 1962; Underwood, 1961; Underwood, Rehula, & Keppel, 1962), so that the theoretical issue of all-or-none vs. gradual association formation remains a matter of considerable dispute.

The present paper reports a series of experiments designed to eliminate or minimize the influence of several inadequately controlled variables in the procedure originally introduced by Rock (1957), wherein performance of a repetition group receiving the same list of paired-associate items on each trial did not differ from a modified nonrepetition group, for whom all pairs not correctly responded to on any trial were removed from the list and replaced by new pairs on the

following trial. Since groups learning under the repetition and nonrepetition procedures may also differ in such factors as instructions and approach to the task (Brackett, 1961), or amount of interference produced by other pairs in the list (Brown, 1961; Clark et al., 1960), a within-$S$s design was employed so that each $S$ learned simultaneously under both conditions and therefore served as his own control. To test more adequately the role of repetition in association formation as distinguished from other processes involved in paired-associate learning (e.g., Underwood & Schulz, 1960), an attempt was made to develop materials which provided a relatively pure and substantial case of association formation. These materials were also highly homogeneous and carefully calibrated so as to be equivalent in difficulty, in order to minimize the effects of selective elimination of the more difficult incorrect pairs on each trial, which has been shown to result in a significantly easier list under the nonrepetition condition (Underwood et al., 1962; Williams, 1961).

## EXPERIMENTS I AND II

### Method

The method and procedure for both Exp. I and II closely followed Rock (1957), except that the list for each $S$ consisted of two equal-sized subsets of pairs representing the repetition (R) and nonrepetition (NR) conditions. All pairs of the R subset were presented on every trial throughout the experiment, whereas only those pairs responded to correctly on the preceding trial were retained in the NR subset, all incorrect pairs being removed from the list and replaced on the next trial

by new pairs. The 15 Ss in each experiment were given typical paired-associate instructions with an added statement that new pairs might be introduced during the experiment, although the procedure effectively prevented S from distinguishing between the R and NR subsets, or from detecting the basis for introducing new NR pairs.

All learning materials were typed on 3×5 in. cards and presented manually through a card-exposure device. Each trial began with the successive presentation of all pairs in the list for 3 sec. each with a 5-sec. interpair interval, followed by all stimulus terms alone in a different order at a 5-sec. rate while S attempted to recall the associated response, with a 30-sec. intertrial interval. To further guard against differentiation of the R and NR subsets, both were distributed evenly throughout the list on each trial.

Experiments I and II differed only in the kind of learning material employed. In Exp. I, 84 pairs of nonsense syllables of 47–53% association value (Glaze, 1928) were selected so as to be maximally homogeneous in judged ease of learning according to a previous scaling procedure (Battig, 1959), and divided into 14 subsets of 6 pairs, matched as to mean and SD of these scale values. The initial list of 12 pairs for each S consisted of two of these equivalent 6-pair subsets representing the R and NR conditions, replacement pairs for the latter being selected from another equivalent subset until exhausted, then from another subset, and so on until the final trial. Each of the 14 subsets was used equally often as the R and initial NR subset and at each point in the sequence of replacement pairs for the various Ss. Learning was carried out to a criterion of 1 errorless trial or a maximum of 10 trials.

Pairs of common 4-letter words as stimuli and 2-digit number responses were used in Exp. II, wherein each S learned to a criterion of one errorless trial an 18-pair list consisting of two 9-pair subsets of R and NR pairs. A total of 84 such word-number pairs were constructed so as to avoid obvious intrapair associations or similarities, each being used equally often in the R and initial NR subset and at various points in the sequence of NR replacements.

Each S in both Exp. I and II was given a recognition test immediately after the last trial, consisting of the individual presentation of 10 incorrect NR pairs which had been replaced, intermixed randomly with 10 additional pairs from the pool of 84 which S had never seen, with S indicating in each case whether or not he thought he had seen the pair previously during the experiment.

## Results

Means and SDs for the repetition (R) and nonrepetition (NR) conditions are presented separately in Table 1 for total errors over the six trials on which data were available from all 15 Ss of Exp. I, and for total errors and trials to the criterion of one errorless trial in Exp. II. Revealed herein is a marked and statistically significant difference in favor of Cond. R in Exp. I ($t = 6.44$, $df = 14$), whereas the slight superiority of Cond. NR for both measures in Exp. II fell far short of significance (both $t$'s < 1). However, the recognition test showed significantly above-chance recognition of incorrect NR pairs which was actually superior in Exp. II (79.7% correct, $t = 7.76$) to Exp. I (70% correct, $t = 7.75$), demonstrating that something had been learned about the replaced incorrect pairs even in Exp. II where their removal did not retard NR learning.

The results of Exp. I, wherein every 1 of the 15 Ss made more NR than R errors, may be due to facilitation by repetition of the substantial response learning required for the nonsense-syllable materials, and therefore are inadequate to disprove an all-or-none theory of association formation as distinguished from paired-associate learning (e.g., Underwood et al., 1962). Although such response learning was minimized for the word-number pairs of Exp. II, so also was the amount of association formation significantly reduced, resulting in such rapid learning (63% correct responses on Trial 2) that the performance measures become quite insensitive to any effects of repetition. Since these materials also differed considerably in difficulty, thereby providing an important advantage for Cond. NR, the results of Exp. II cannot be interpreted as unequivocally supporting an all-or-none theory of association formation.

## Experiment III

To overcome the major deficiencies of Exp. I and II, an attempt was made to develop for use in Exp. III materials which required a substantial amount of association formation while relatively free from other nonassociative processes typically involved in paired-associate learning. This was accomplished by (a) constructing pairs with maximally different and unrelated stimulus and response members; (b) using stimuli of low meaningfulness and familiarity which were highly discriminable from each other, since previous evidence indicates that discriminability represents a major factor in learning of the stimulus (Battig, Williams, & Williams, 1962), whereas meaningfulness and familiarity are of lesser importance (Underwood & Schulz, 1960); (c) using highly meaningful and familiar responses to minimize response learning.

### Method

Each of 72 nonsense shapes of high discriminability and low association value was paired with a different two-digit number between 12 and 98 (excluding double numbers and numbers ending in zero), forming 72 pairs with minimal shape-number similarity according to the judgments of 10 preliminary Ss. Previous measures had been obtained of (a) association value and discriminability of each shape from 122 Ss; (b) association value of each number from 28 Ss; (c) rated "ease of learning" of each shape-number pair by 66 Ss; (d) actual learning difficulty in a separate study reported elsewhere, which describes the materials in more detail (Battig & Brackett, 1961). These 72 pairs were divided into 12 subsets of 6 pairs each which were approximately equivalent with respect to each of these four indices, each being used with approximately equal frequency as the R and initial NR subset and throughout the sequence of NR replacement pairs.

All other aspects of method and procedure were identical to Exp. I with the following minor exceptions: (a) all materials were photographed and presented for learning and attempted recall by means of an automatic slide projector; (b) a 5-sec. rate (4-sec. ex-

### TABLE 1

Performance Measures for Repetition and Nonrepetition Conditions of Exp. I–III

| Exp. | Measure | Repetition | | Nonrepetition | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| I | Errors | 22.13 | 5.40 | 28.93 | 5.25 |
| II | Errors | 13.27 | 6.95 | 12.40 | 5.87 |
| | Trials | 4.47 | 1.41 | 4.20 | 1.17 |
| III | Errors | 16.00 | 4.87 | 20.07 | 7.04 |
| | Trials | 5.80 | 1.17 | 6.67 | 1.85 |

Note.—Errors and trials are to a criterion of one errorless trial, except for Exp. I which is based on Trials 1–6 for all Ss.

posure, 1-sec. interslide interval) was used for presentation of pairs for learning as well as for stimuli alone during attempted recall, with a 30-sec. interval between trials and between presentation and recall series within each trial; (c) learning was continued to a criterion of two successive errorless trials (although none of the 15 Ss ever made an error on the trial following the first errorless trial); (d) following the postexperimental recognition test for incorrect NR pairs, each S was shown again each of the 12 pairs he had learned and asked to describe verbally how he had learned them.

### Results

As shown in Table 1, performance was superior on the R subset according to both error and trial measures, although significantly so only for errors ($t = 2.83$, $P < .02$) but not for trials ($t = 1.85$, $.05 < P < .10$). Further error analysis revealed the difference to be primarily due to significantly more response omissions under Cond. NR ($t = 3.06$, $P < .01$), which accounted for 29% of the NR errors as compared with 18% for Cond. R. The two conditions did not differ significantly either in partially correct responses ($t = 1.25$), intralist intrusions ($t = 1.34$), or extralist intrusions ($t = 1.05$). Recognition test results were similar to Exp. I and II, yielding 71.7% correct identifications

EXPERIMENT IV
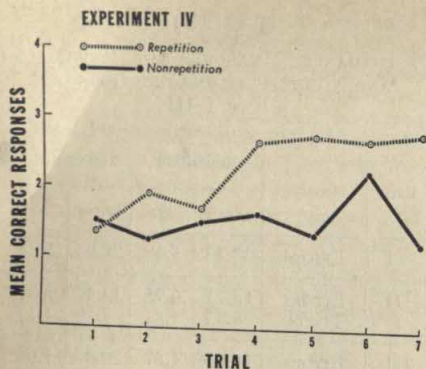


FIG. 1. Mean correct responses per trial for the modified repetition and nonrepetition conditions of Exp. IV on Trials 1–7.

($t = 9.12$). Further analysis revealed recognition errors to consist primarily of reports of incorrect NR pairs as not seen previously (85%) rather than of the unseen pairs as previously seen (15%).

Based on postexperimental questioning, the learned pairs were categorized on the basis of $S$'s ability to verbalize a mediated basis for associating the shape and number together, yielding insignificantly more mediated associations for NR (24%) than R pairs (20%). No basis for association could be verbalized for 67% and 63% of the R and NR pairs, respectively, the remaining 13% in each case representing a "doubtful" category. These results indicate that the present materials had been reasonably successful in providing a relatively pure and substantial case of association formation, and were sufficiently homogeneous to minimize if not eliminate the problem of differences in pair difficulty in favor of Cond. NR.

### EXPERIMENT IV

Although the differences in favor of R over NR performance in Exp. III were significant statistically, they were not impressively large in magnitude, due at least in part to the relatively small number of pairs learned by each $S$, and the predominance of already learned pairs on all but the first few trials as a consequence of the rapid learning of the shape-number pairs. In order to overcome these deficiencies and provide a more sensitive test of the role of repetition in this task, the procedure was modified in Exp. IV so that all pairs learned on each trial were eliminated from the list and replaced by new pairs.

### Method

As in Exp. III, each $S$ was presented with a list of 12 shape-number pairs on each trial, divided into two equivalent subsets of 6 pairs representing repetition (R) and nonrepetition (NR) conditions. However, all NR pairs were presented only for a single trial, being replaced after each trial by another of the equivalent subsets of 6 pairs which $S$ had not seen previously. Pairs of the R subset were removed from the list and replaced by new pairs only if they had been responded to correctly on the preceding trial, so that this subset on any given trial typically included a combination of new pairs with previously presented incorrect pairs. Trials were continued in this manner until the pool of 72 pairs was exhausted, which required seven–nine trials depending on the number of correct responses to R pairs for the various $S$s.

The only other differences in method and procedure from Exp. III consisted of (a) appropriate modifications in instructions to cover the changed conditions; (b) elimination of the postexperimental recognition test and attempted verbalization of correct associations; (c) use of 20 paid $S$s representing a wider range of previous experience and sophistication about paired-associate learning experiments than in Exp. I–III, each of which had used 15 volunteers from undergraduate psychology courses.

### Results

The mean number of correct responses per trial for pairs of the R and NR subsets are presented in Fig. 1, which indicates a consistent and increasing superiority of Cond. R to NR over the seven trials on which data were available from all $S$s.

Total correct responses for Cond. R (Mean = 17.85, $SD$ = 4.24) were significantly greater ($t$ = 7.06, $df$ = 19) than for Cond. NR (Mean = 11.65, $SD$ = 4.96). Except for 2 $Ss$ who performed identically under both conditions, all $Ss$ made more correct responses on the R subset. The slopes of the R and NR curves of Fig. 1 were found to differ significantly by trend analysis of variance ($F$ = 10.35, $df$ = 1/38), due to a highly significant improvement over trials for Cond. R ($F$ = 29.92, $df$ = 1/19) whereas the slight increase in correct responses under Cond. NR fell far short of significance ($F < 1$). In agreement with Exp. III, error analysis revealed significantly more omissions ($t$ = 2.86, $P < .01$) for Cond. NR (Mean = 11.10) than Cond. R (Mean = 7.50), but the excess of partially correct responses and intrusions in Cond. NR (Mean = 24.35) over Cond. R (Mean = 21.75) was not significant ($t$ = 1.58, $P > .10$).

Although the greater increase in correct responses over trials under Cond. R undoubtedly reflects the increased proportion of frequently presented incorrect pairs on later trials, a more direct and convincing demonstration of the superiority of performance on repeated pairs comes from a comparison of proportions of correct responses for R pairs varying

in number of prior presentations. Table 2 presents the total number of pairs presented once, twice, etc. within the R subset summed over all $Ss$, along with the total number and proportion of these presentations which yielded correct responses, and the mean proportion of correct responses averaged over $Ss$[2]. Revealed herein is a consistent increase with number of prior presentations in both total and mean proportion of correct responses, the rate of increase being slightly less for total proportion due to the relatively greater contribution of slower learners to this figure as number of presentations increase. The proportions of first, second, and third presentations responded to correctly by each $S$ were subjected to trend analysis of variance, which showed the increase with number of presentations to be highly significant ($F$ = 24.56, $df$ = 1/19).

### Discussion

The present results demonstrate conclusively the facilitation of paired-associate learning by repetition of incorrect pairs under conditions where ($a$) other variables which may differ for R and NR conditions have been eliminated through the use of a within-$Ss$ design; ($b$) careful precautions have been taken to minimize the effects of differences between pairs in difficulty; ($c$) nonassociative factors in paired-associate learning, such as response learning

TABLE 2

Proportions of Correct Responses for Repetition Pairs as a Function of Number of Prior Presentations

| Measure | Number of Prior Presentations | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5+ |
| Total number | 420 | 265 | 135 | 66 | 56 |
| Correct responses | 118 | 114 | 61 | 32 | 28 |
| Total proportion | .281 | .430 | .452 | .485 | .500 |
| Mean proportion | .270 | .424 | .496 | .581 | .539 |

[2] A somewhat more stable estimate of performance on once-presented pairs would probably be provided by the inclusion of pairs of the NR subset, all of which were presented only once. However, performance on NR pairs ($p$ = .247) was somewhat below that for R pairs presented for the first time ($p$ = .281), probably due to the differential distribution of these pairs over trials under the two conditions. The decision to base the present comparison solely on R pairs therefore represents, if anything, an overly conservative estimate of the magnitude of the increase in proportion of correct responses with number of previous presentations.

and stimulus discrimination, have been largely eliminated from the task. Only in Exp. II, which was clearly inadequate due to the extremely rapid learning of the word-number pairs and the lack of control for differences in pair difficulty, were the results not in direct conflict with an all-or-none theory, and even here the results of the recognition test showed that something short of a correct association had been learned about the eliminated incorrect NR pairs.

In comparison with previous studies using the Rock (1957) procedure, Exp. III and IV would appear to be clearly superior with respect to the elimination of uncontrolled variables and biases favoring either the R or NR condition. However, despite the extensive efforts to equate the shape-number pairs in average difficulty, these pairs probably were not equally difficult for the individual S. Furthermore, besides being more difficult, the increasing superiority of performance on more frequently presented R pairs in Exp. IV may have obtained in spite of the relatively large contribution of slow learners to measures based on these pairs. Since either or both of these sources of bias would clearly favor NR performance, the conclusiveness of the present results in demonstrating the positive role of repetition can only be enhanced thereby.

Only if the position is taken that some process other than association formation has been facilitated by repetition can the results of Exp. III and IV be regarded as insufficient to disprove an all-or-none theory of association formation. Any such argument, of course, reduces the issue to empirical meaninglessness unless the association-formation process can be given precise operational specification. While verbal definitions of association formation as distinguished from paired-associate learning suffer from considerable ambiguity, there seems to be reasonable agreement that this refers to the process of "connecting," "hooking-up," or "associating together" the stimulus and response members of each pair. Procedurally, efforts to provide the necessary pure case of association forma-

tion have concentrated on the elimination of response learning, reflecting the influence of Underwood's conception of paired-associate learning as a two-stage process consisting of response-learning and associative phases (Underwood et al., 1962; Underwood, Runquist, & Schulz, 1959; Underwood & Schulz, 1960). There can be little question but that the present results are directly applicable to this definition of association formation as paired-associate learning in the absence of response learning.

However, although the present stimulus materials were designed to be maximally discriminable, thereby eliminating another nonassociative process suggested to be involved in paired-associate learning (Battig et al., 1962; Gibson, 1940), they differed from the letters or words typically employed as stimuli in related studies in their low meaningfulness and familiarity. Inasmuch as stimulus familiarity has little or no effect on paired-associate learning while stimulus meaningfulness affects primarily the association-formation phase (Underwood & Schulz, 1960), the present shape-number materials would therefore appear to be maximally appropriate and sensitive as a test of the role of repetition in association formation, since they do not substantially reduce, nor shortcut through already existing mediated associations, the amount of association formation required of S.

Nevertheless, although the present materials have eliminated those nonassociative processes previously identified to be important in paired-associate learning, the possibility remains that other processes besides association formation are still involved in learning the shape-number pairs. However, unless or until such processes are identified and given adequate operational specification, it can be concluded that association formation as presently defined is not an all-or-none process, but instead builds up gradually in strength through repetition. Moreover, in view of the rapid learning observed even under the present conditions, which were carefully designed to maximize the required amount of

association formation, it would appear that this process may constitute such a small and insignificant part of the learning involved in the typical paired-associate task, that questions concerning its all-or-none or gradual nature are likely to be of little consequence for the general understanding of factors important in paired-associate learning.

## SUMMARY

The effect of repetition of previously incorrect pairs in paired-associate learning was evaluated in four separate experiments in which each S learned a single paired-associate list consisting of two equivalent subsets of pairs. In Exp. I–III, each of 15 Ss learned under conditions where the repetition (R) subset consisted of the same pairs repeated on all trials, while pairs of the nonrepetition (NR) subset were retained in the list only if responded to correctly and were otherwise removed from the list and replaced by new pairs on the next trial. Lists of 12 nonsense-syllable pairs, 18 word-number pairs, and 12 pairs of nonsense shapes and numbers were used in Exp. I–III, respectively. Experiment IV ($N = 20$) used the same shape-number pairs as Exp. III, but under conditions where all pairs responded to correctly were immediately removed from the list, so that the NR subset consisted of a new set of pairs on each trial, while the R subset included both new pairs and previously presented incorrect pairs. Significant differences in favor of Cond. R were obtained in all cases except for Exp. II, which was attributed to the greater ease of learning and rather wide range of difficulty of the word-number pairs. Particularly in view of the significant positive effects of repetition in Exp. III and IV, using materials requiring a relatively pure and substantial case of association formation which were also carefully calibrated to minimize the effects of differences in pair difficulty, it was concluded that the present results are directly contradictory to an all-or-none theory of association formation in paired-associate learning.

## REFERENCES

BATTIG, W. F. Scaled difficulty of nonsense-syllable pairs consisting of syllables of equal association value. *Psychol. Rep.*, 1959, **5**, 126.

BATTIG, W. F., & BRACKETT, H. R. Comparison of anticipation and recall methods of paired-associate learning. *Psychol. Rep.*, 1961, **9**, 59–65.

BATTIG, W. F., WILLIAMS, J. M., & WILLIAMS, J. G. Transfer from verbal-discrimination to paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 258–268.

BRACKETT, H. R. The effect of pretraining and knowledge of results on the acquisition of paired associates. Unpublished master's thesis, University of Virginia, 1961.

BROWN, S. C. Interpair interference in paired-associate learning. Unpublished master's thesis, University of Virginia, 1961.

CLARK, L. L., LANSFORD, T. G., & DALLENBACH, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73**, 22–40.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

POSTMAN, L. Repetition and paired-associate learning. *Amer. J. Psychol.*, 1962, **75**, in press.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

ROCK, I., & HEIMER, W. Further evidence of one-trial associative learning. *Amer. J. Psychol.*, 1959, **72**, 1–16.

UNDERWOOD, B. J. One-trial learning. Invited address at Midwestern Psychological Association, Chicago, May 1961.

UNDERWOOD, B. J., REHULA, R., & KEPPEL, G. Item selection in paired-associate learning. *Amer. J. Psychol.*, 1962, **75**, in press.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intra-list similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

WILLIAMS, J. P. Supplementary report: A selection artifact in Rock's study of the role of repetition. *J. exp. Psychol.*, 1961, **62**, 627–628.

# SUPPLEMENTARY REPORT: EFFECTS OF STIMULUS ASSOCIATION VALUE AND EXPOSURE DURATION ON R-S LEARNING [1]

NED CASSEM AND DONALD H. KAUSLER

*St. Louis University*

Jantz and Underwood (1958) demonstrated a gross positive relationship between stimulus association value (AV) and the amount of R-S learning. However, they found the same reversal between 0% and 33% Glaze values that had been reported earlier by Postman, Adams, and Phillips (1955). As noted by Jantz and Underwood, the reversal most likely represented a sampling artifact in that only two syllables were included in their list at each AV level. The present study attempted to provide a more precise determination of the relationship between the AV of stimuli and R-S learning by means of a more adequate sampling of stimulus items.

In addition, the present study investigated the effects of exposure duration during both S-R learning and R-S recall on the amount of R-S recall. If, as proposed by Feldman and Underwood (1957), R-S learning is a variant of incidental learning, in which *S* learns associations for which he has no learning set during a period in which he is set to learn other associations, then the amount of R-S or incidental learning should increase with increased exposure duration (Kausler & Trapp, 1961; Rosenberg, 1959).

*Method.*—Eighty seminarians were randomly assigned to the 16 groups of a $2 \times 2 \times 4$ factorial design, representing exposure duration during S-R learning ($T_L$), exposure duration during R-S recall ($T_R$), and AV. Each *S* had six trials on an eight-item list with nonsense syllables as stimuli and the same words employed by Jantz and Underwood as responses. Four sets of eight syllables were selected from the 0, 33, 67, and 100% Glaze lists, with the additional criterion that the values approximate Archer's (1960) revaluation. Within each set similarity was minimized and four different combinations of S-R pairings were used to control for associative effects. Both $T_L$ and $T_R$ were either 2 sec. or 4 sec. In other respects the experiment replicated that of Jantz and Underwood (1958).

*Results.*—As found in the earlier study, total syllables correct during the recall trial displayed significant heterogeneity of vari-

ance between groups. The total letters correct, a criterion used by Feldman and Underwood (1957), indicated homogeneity of variance. Consequently, the latter measure served in the analysis of variance of treatment effects reported here. However, comparable results were obtained throughout for the heterogeneous total syllable scores.

The overall means and *SD*s for AV were 10.30 and 4.77, 14.70 and 5.29, 16.60 and 4.16, and 19.40 and 3.82 for 0, 33, 67, and 100%, respectively. A trend analysis for these means revealed a highly significant linear trend ($F = 22.34$, $P < .001$). The overall means and *SD*s for $T_L$ were 12.52 and 5.17 for 2 sec. and 18.20 and 3.92 for 4 sec. The corresponding $T_R$ statistics were 13.90 and 4.54 for 2 sec. and 16.83 and 4.55 for 4 sec. The main effects for AV, $T_L$, and $T_R$ were all significant ($F = 10.85$, $df = 3/64$, $P < .001$; $F = 26.98$, $df = 1/64$, $P < .001$; $F = 7.14$, $df = 1/64$, $P < .01$). Of the interactions, only the $T_L \times T_R$ approached significance ($F = 2.85$, $df = 1/64$, $P < .10$). This trend suggested that $T_R$ had little effect on the 2-sec. $T_L$ groups but did on the 4-sec. $T_L$ groups. That is, the 4-sec. recall condition was most effective in combination with the 4-sec. S-R learning condition.

The present results appear to verify the assumption made by Jantz and Underwood (1958) that their reversal for 0% and 33% represented a sampling artifact. The results also indicate that R-S learning is sensitive to exposure duration during both the learning and recall periods.

## REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

KAUSLER, D. H., & TRAPP, E. P. The effects of position and exposure duration on irrelevant cue learning. Paper read at Southwestern Psychological Association, Little Rock, Arkansas, April 1961.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

ROSENBERG, S. Exposure interval in incidental learning. *Psychol. Rep.*, 1959, **5**, 675.

# SUPPLEMENTARY REPORT: EFFECTS OF INSTRUCTIONS ON EXTINCTION AND RECOVERY OF A CONDITIONED AVOIDANCE RESPONSE

K. E. MOYER                                       AND                    RICHARD H. LINDLEY[1]

*Carnegie Institute of Technology*                                   *Trinity University, Texas*

Lindley and Moyer (1961) found that, in the case of a classically conditioned finger withdrawal response, informing *S*s just prior to extinction training that the UCS would no longer be delivered produced rapid extinction. The present study replicated that study, but with a conditioned avoidance procedure, and in addition tested an implication of the drive level (*D*) interpretation of their result; viz., that the appropriate instructions could raise *D* level and lead to an increase in conditioned responding.

*Method.*—The same experimental room, apparatus, intertrial stimulus intervals, method of adjusting shock level, and pretests with the tone alone were used (Lindley & Moyer, 1961). The CS was a .5-sec. tone which preceded a 1.5-sec. electric shock without overlap. The finger could only move up when the CS and UCS were presented. No shock was delivered if *S* responded to the CS and kept his finger raised $\frac{1}{8}$ in. or more during the UCS interval. If *S* did not avoid the shock, moving the finger up $\frac{1}{8}$ in. escaped the shock.

The booth in which *S* sat was dark. The initial instructions were similar to the previous instructions except that the avoidance rather than the classical CR procedure was explained. The *S* was told that we wanted to condition automatic movements of the finger but that he should not try to prevent his finger from moving.

A factorial design involving number of acquisition conditioning trials (to four CRs in 5 consecutive trials, or that plus 20 trials), instructions prior to extinction (Neutral or Inhibitory), and instructions after 20 extinction trials (No Instructions or Resume UCS) was used. The Neutral instructions prior to extinction were: "Be sure to let your finger jump up to the tone when it feels like it"; the Inhibitory instructions were: "There will be no more shock presented from now on. I want you to try to prevent your finger from moving when the tone is presented." After 20 extinction trials half the *S*s were given an additional 5 extinction trials without any instructions (No Instructions); the remaining *S*s were told: "From now on the tone will be followed by shock on most of the trials. Remember to let your finger jump up to the

tone if it feels like it" (Rseume UCS). The extinction procedure was continued for 5 trials (i.e., no shock was delivered). The normal intertrial interval was used between Trials 20 and 21 for the No Instructions groups; this interval was longer for the Resume UCS groups owing to the reading of the instructions.

The 56 *S*s (both men and women) from Carnegie Institute of Technology who reached criterion within 35 trials were assigned to the eight groups (*N* per group = 7) by a randomized block procedure. Twenty-seven *S*s were rejected owing to failure to reach criterion within 35 trials, apparatus failure, or unwillingness to be shocked.

*Results.*—The mean number of trials to reach the criterion of four CRs in five consecutive trials was 10.98, *SD* = 7.16. There were no significant differences among the groups in reaching this criterion (*F* = 0.30).

In extinction, a CR was defined as any response that occurred during the .5-sec.
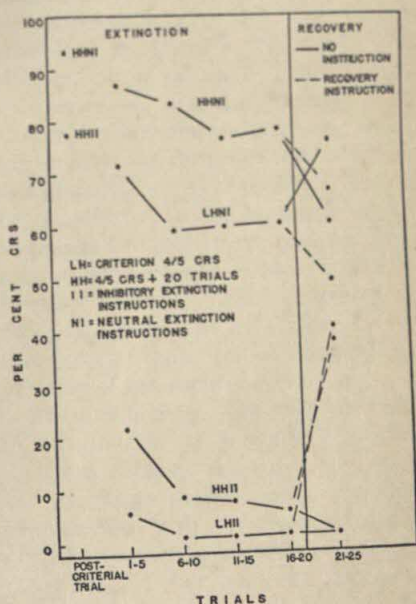


FIG. 1. Percentage of CRs for blocks of five trials during extinction and recovery training and the percentage of CRs on the first postcriterial trial for the 4/5 + 20 trial groups.

duration of the CS. The variances of the number of CRs in extinction were highly heterogeneous ($x^2 = 62.53$, $df = 3$, $P < .001$). Owing to the heterogeneity of variance, the data of the groups that received neutral instructions and of the groups that received inhibitory instructions were analyzed separately and the $x^2$ median test was used at times instead of the more usual $t$ or $F$ tests.

A $x^2$ median test showed that the inhibitory instructions significantly reduced the number of CRs elicited in the 20 extinction trials as compared to the neutral instructions ($x^2 = 23.17$, $df = 1$, $P < .001$). Figure 1 shows the extinction and recovery data for all groups and also the percentage of CRs on the first trial after $S$s reached the $\frac{4}{5}$ criterion in the two groups with the $\frac{4}{5}$ criterion plus 20 trials. The groups with the additional 20 conditioning trials gave more responses in extinction than the groups that merely met the $\frac{4}{5}$ criterion (for the two groups with Neutral instructions, $t = 1.27$, $df = 26$, $P > .05$; for the groups with Inhibitory instructions, $t = 2.18$, $df = 26$, $P < .05$).

The instructions designed to raise $D$ level after 20 extinction trials had a significant effect on the groups that received Inhibitory instructions in extinction (median test $x^2 = 8.58$, $df = 1$, $P < .01$). The same instructions had no demonstrable effect on the groups that received Neutral instructions in extinction ($x^2 = 1.29$, $df = 1$, $P > .05$).

As in classical conditioning, inhibitory instructions dramatically reduced the number of CRs during extinction after avoidance conditioning, especially after the larger number of conditioning trials. This fits in with the notion that a habit factor is built up through reinforcement and is independent of a performance or drive factor which can be manipulated by instructions. The results also indicate that, for $S$s who have been given inhibitory instructions, the appropriate instructions can increase the number of CRs after extinction, presumably by raising $D$ level. The latter result was not found for $S$s who have been given neutral instructions in extinction, perhaps due to the already high level of responding of these groups after the 20 extinction trials.

### REFERENCE

LINDLEY, R. H., & MOYER, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, 61, 82–88.

## SECONDARY REINFORCEMENT IN RATS AS A FUNCTION OF INFORMATION VALUE AND RELIABILITY OF THE STIMULUS [1]

M. DAVID EGGER AND NEAL E. MILLER

*Yale University*

Although secondary reinforcement has been of major importance to behavior theory, especially in explanations of complex learning phenomena (e.g., Hull, 1943; Miller, 1951; Skinner, 1938), little is known about the conditions for its occurrence in any but the simplest situations. The first hypothesis explored in the experiments reported here is that in a situation in which there is more than one stimulus predicting primary reinforcement, e.g., food, the more informative stimulus will be the more effective secondary reinforcer. Further it is asserted that a necessary condition for establishing any stimulus as a secondary reinforcer is that the stimulus provide information about the occurrence of primary reinforcement; a redundant predictor of primary reinforcement should not acquire secondary reinforcement strength.

A possible situation in which to test this hypothesis is the following: a short stimulus always precedes the delivery of food. But it is made essentially redundant by being overlapped by a longer stimulus of slightly earlier onset which is also invariably followed by food. This situation is summarized in Fig. 1. The longer stimulus is labeled $S_1$ and the shorter, $S_2$. For an $S$ trained with this series of stimulus events, $S_2$ is a reliable, but redundant, i.e., noninformative, predictor of food. Hence, according to our hypothesis, $S_1$ should be an effective secondary reinforcer; $S_2$ should acquire little or no secondary reinforcing strength, even though it is closer in time to the occurrence of food, and therefore in a more favorable position than is $S_1$ on the gradient of delay of reinforcement.

There is a way, however, to make $S_2$ informative. If $S_1$ occurs a number of times without $S_2$, unaccompanied by the food pellet, and randomly interspersed with occurrences of the stimulus sequence shown at the bottom of Fig. 1, then $S_2$, when it occurs, is no longer redundant; for now $S_2$ is the only reliable predictor of food. Thus, it is predicted that for a group

97
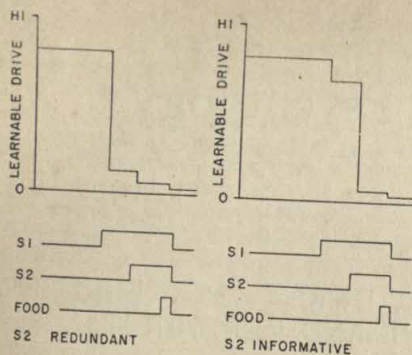
FIG. 1. Schematic representation of the theoretical analysis of the two Main Experiment groups according to a strict interpretation of the drive-reduction hypothesis.

of rats who receive the stimulus sequence depicted in Fig. 1 interspersed with occurrences of $S_1$ alone, $S_2$ will be a considerably more effective secondary reinforcer than for the group of rats who receive only the stimulus sequence depicted in Fig. 1.

It should be noted that both groups will receive exactly the same number of pairings of $S_2$ with food and in exactly the same immediate stimulus context, so that if a difference were found between the groups in the secondary reinforcing value of $S_2$, it could not be due to simple patterning, stimulus-generalization decrement, or differences in association with food.

Our predicted results would be compatible with a strict interpretation of the drive-reduction hypothesis of reinforcement (Miller, 1959). Such a theoretical analysis is represented schematically in the upper portion of Fig. 1. According to the drive-reduction hypothesis, a stimulus acquires secondary reinforcing value by acquiring the ability to elicit a drive-reducing response. The left side of Fig. 1 illustrates that if most of the learnable drive already has been reduced by $S_1$, little drive-reduction remains to be conditioned to $S_2$.

On the other hand, if $S_1$ sometimes fails to predict food, some of the conditioned drive-reduction to it should extinguish. Hence, as is depicted on the right side of Fig. 1, more of the drive-reduction should occur to, and be conditioned to, $S_2$.

From Fig. 1, one can also see that the drive-reduction analysis also demands that the secondary reinforcing value of $S_1$ should be greater when it is a reliable predictor (making $S_2$ redundant) than when it is an unreliable predictor (making $S_2$ informative). Thus we are led to our second hypothesis, namely, that in a situation in which a predictor of primary reinforcement exists which is both reliable and informative, this predictor should become a more effective secondary reinforcer than an unreliable predictor. Note that here we predict the opposite of a partial-reinforcement effect, which would be expected to increase the resistance to extinction of the unreliable predictor, that is, the stimulus which had been paired with food only part of the time. In any prolonged test for secondary reinforcement, this increased resistance to extinction should show up as a greater total secondary-reinforcing effect.

## MAIN EXPERIMENT
### Method

*Subjects.*—The Ss were 88 male rats of the Sprague-Dawley strain who were approximately 90 days old at the beginning of their experimental training. Owing to deaths and equipment failures, the data from 4 Ss were lost, and the data from another 4, selected at random, were discarded in order to have equal sized groups for an analysis of variance. The Ss, fed once daily following the experimental session, were maintained at approximately 80% of their ad lib. weight.

*Apparatus.*—The apparatus consisted of two identical Skinner boxes, 19 in. long, 8 in. wide, and 6¾ in. high (inside dimensions). The floors of the boxes consisted of six ½ in.-

diameter rods running parallel to the side containing the Plexiglas door. Each box was enclosed in a large, light-proof, sound-deadened crate into which a stream of air was piped for ventilation and masking noise. Inside each of the Skinner boxes were two lights, one located 2 in. above the food cup, another located in the middle of the long back wall, opposite the Plexiglas door. The food cup was in the center of the front, 8-in. wall; the bar, a bent steel strip $1\frac{1}{2}$ in. wide, protruded $\frac{1}{2}$ in. into the inner chamber of the box. The entire bar assembly was removable and, when withdrawn, its opening was sealed with a metal panel. The bar was located to the right of and slightly above the food cup. A downward force of at least 12 gm. on the bar activated a microswitch normally connected in the circuit of a Gerbrands feeder which delivered a standard .045-gm. Noyes pellet into the food cup. A loudspeaker was located 3 in. behind and slightly to the left of the front wall of the Skinner box. Both flashing lights (12 per sec.) and tones (750 cps) were used as stimuli.

*Procedure.*—All training sessions lasted 25 min. per day. During the first three sessions, Ss were magazine-trained in the absence of the bar. Then the bar was inserted, and, for two sessions, each bar press was followed by a pellet of food. A few rats who did not spontaneously learn to press were given an extra remedial session during which bar pressing was "shaped." Over the next four sessions the required ratio of responses to reinforcements was gradually increased to 4:1.

Then, for the subsequent five sessions, the bar was removed, and Ss were randomly assigned to Group A (for whom S₂ was reliable but *redundant*) and Group B (for whom S₂ was reliable and *informative*). Group A received the following sequence of events during each of its five "stimulus-training" sessions: once every 56 sec. on the average, a pellet of food was delivered into the food cup. The pellet was inevitably preceded by 2 sec. of $S_1$ and $1\frac{1}{2}$ sec. of $S_2$. Both stimuli overlapped the delivery of the food pellet by $\frac{1}{4}$ sec., and both terminated together.

Group B also received this stimulus sequence immediately preceding the delivery of the food pellet. But in addition, Group B Ss received aperiodically, interspersed with the stimulus-food sequence, 2 sec. of $S_1$ alone. The events for Group B occurred on the average of once every 30 sec.

For half the Ss in each group, $S_1$ was a flashing light and $S_2$ was a tone, and for the other half, the conditions were reversed: $S_1$ was a tone and $S_2$ was a flashing light.

During 5 days of such training, each group received 135 pairings of $S_1$ and $S_2$ with food, and Group B received in addition about 110 occurrences of $S_1$ alone. Thus for both groups $S_2$ was followed 100% of the time by food, while $S_1$ was followed by food 100% of the time for Group A, but only 55% of the time for Group B.

The above description of training applies to all but 16 Ss, 8 Group B and 8 Group A. For these Ss, training was exactly as described above except that the stimulus-food pairings occurred for both groups on the average of once every 75 sec. instead of 56 sec., and Group B received a stimulus event on the average of once every 15 sec. instead of 30 sec., so that $S_1$ was followed by food only 20% of the time for Group B. These 16 Ss were given seven 25-min. "stimulus-training" sessions. The data from these Ss were analyzed separately and not included in the overall analysis of variance.

*Testing.*—On the day following the final stimulus-training session, Ss were tested as follows: the bar was reinserted and Test Session 1 began with each S pressing for food pellets on a fixed ratio of 3:1. The retraining presses continued until S had received 30 pellets. At this point the bar was disconnected and 10 min. of extinction ensued.

At the end of the 10 min., the bar was reconnected, not to the feeder, but to a timer which delivered on the same 3:1 schedule 1 sec. of whatever stimulus was being tested for secondary reinforcing strength. The test session continued until 25 min. had elapsed since the beginning of the extinction period, or until 10 min. after the first occurrence of a stimulus, whichever was longer.

In the foregoing procedure, relearning following experimental extinction was used as the measure of secondary reinforcing strength on the assumption that it would be more rapid and less variable than would de novo learning of the skill of pressing the bar. A preliminary study had validated this technique showing that in such a test more bar presses would occur when followed by a stimulus previously associated with food than when the stimulus had not been associated with food.

After an interval of 1 day, Test Session 2 was conducted, identical to the first, except that this time the stimulus delivered following the 10-min. extinction period was the opposite from that tested in Test Session 1: for half

of the Ss, $S_2$ was tested in Test Session 1 and $S_1$ was tested in Session 2; for the other half of the Ss, trained and tested subsequent to the first half, the stimuli were tested in the opposite order.

For Ss tested first with $S_2$ and then with $S_1$, Test Session 3 followed another intervening day, this time with Ss pressing for $S_2$ again. Throughout the course of the 10-min. extinction and ensuing "pressing for stimuli" period, the cumulative total number of bar presses for each S was recorded each minute.

*Response measures.*—The total number of bar presses in a 10-min. period following the first occurrence of the stimulus was the measure of secondary reinforcing strength. Since there were significant between-S and within-S correlations ($r_b = .53$; $r_w = .34$) of this measure with the total number of bar presses in the 10-min. extinction periods, this total number of bar presses in extinction was used as a control variable in analyses of covariance. (It should be noted that most of the bar presses during extinction occurred within the first 2-4 min. of the 10-min. extinction period.)

Furthermore, since it was found that in no case would analyses based only on data from Test Session 1 have led to any substantially different conclusions from those reported below, the means and results of analyses reported (unless otherwise noted) are based on combined data from Test Sessions 1 and 2.

Since by Test Session 3, there no longer appeared to be any differences between the experimental groups, the data from this session were not included in the final analyses.

## Results

*Overall analysis.*—Neither of the hypotheses being tested depended

### TABLE 1

MEAN RESPONSES DURING 10 MIN. OF EXTINCTION AND 10 MIN. OF "PRESSING FOR STIMULI"

| Group | $S_1$ | | $S_2$ | | $S_1 + S_2$ |
|-------|------|----------|------|----------|-------------|
| | Ext. | Pressing | Ext. | Pressing | |
| A | 110.8 | 115.1 | 101.9 | 65.8 | 90.5 |
| B | 112.1 | 76.1 | 112.0 | 82.6 | 79.4 |
| A+B | | 95.6 | | 74.2 | |

Note.—Test Sessions 1 and 2 combined.

upon the significance of the main effects of the overall analysis, but instead upon comparisons between the means shown in specific subcells of Table 1. The marginal entries in Table 1 give the overall means for Groups A and B (rows), and for $S_1$ and $S_2$ (columns). The overall mean for each group is based on data from 32 Ss each tested with $S_1$ and with $S_2$; the overall mean for each stimulus position is based on data from all 64 Ss. As seen from an inspection of Tables 1 and 2, Group A responded significantly more than Group B and the position of $S_1$ was reliably more effective than that of $S_2$.

It should be noted that although the groups were identically treated in all other respects, the 32 Ss tested with $S_1$ first and $S_2$ second were run subsequent to the 32 Ss tested with $S_2$ first and with $S_1$ second. No significant differences between these groups existed in the control variable, total presses in 10 min. of extinction. Nor did an analysis of covariance reveal any significant effects of order of testing (O), or of the interaction of order of testing with experimental group (G), or with stimulus position (P) (see Table 2).

Across all groups, the Ss responded more for the flashing lights than for the tones ($F = 8.45$; $df = 1/55$; $P < .01$, analysis of covariance).

Examination of the minute-by-minute response totals during the "pressing for stimuli" period revealed that the differences between groups tested at 10 min. had generally begun to appear after 3-5 min., and continued to increase out to 15 min., which was the longest period any S was permitted to bar press for stimuli during a given test session.

As expected from our hypotheses, the (P × G) interaction was highly significant ($F = 17.71$; $df = 1/55$;

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE AND COVARIANCE: TEST SESSIONS 1 AND 2 COMBINED

| Source | Analysis of Variance | | | Analysis of Covariance | | |
|---|---|---|---|---|---|---|
| | df | MS | F | df | MS | F |
| **Between $Ss$** | | | | | | |
| Experimental Group (G) | 1 | 3,916.12 | 2.36 | 1 | 6,062.17 | 4.98* |
| Modality of $S_1$ (M) | 1 | 7,938.00 | 4.79* | 1 | 4,792.37 | 3.93 |
| Order of $S_1, S_2$ (O) | 1 | 435.13 | | 1 | 709.44 | |
| G × M | 1 | 2,907.03 | 1.75 | 1 | 573.36 | |
| G × O | 1 | 2,329.03 | 1.41 | 1 | 50.93 | |
| M × O | 1 | 9.03 | | 1 | 37.41 | |
| G × M × O | 1 | 1,624.50 | | 1 | 1,382.64 | 1.14 |
| Error (b) | 56 | 1,657.19 | | 55 | 1,218.04 | ($r_b$ = .53) |
| **Within $Ss$** | | | | | | |
| Stimulus Position (P) | 1 | 14,663.28 | 10.83** | 1 | 12,168.39 | 9.97** |
| P × G | 1 | 24,864.50 | 18.36*** | 1 | 21,613.81 | 17.71*** |
| P × M = St | 1 | 15,664.50 | 11.56** | 1 | 10,316.32 | 8.45** |
| P × O = T | 1 | 52,650.13 | 38.87*** | 1 | 1,594.90 | 1.31 |
| P × G × M | 1 | 87.78 | | 1 | 546.78 | |
| P × M × O | 1 | 35.13 | | 1 | 16.45 | |
| P × G × O | 1 | 5,330.28 | 3.94 | 1 | 3,168.95 | 2.60 |
| P × G × M × O | 1 | 5,781.27 | 4.27* | 1 | 6,720.79 | 5.51* |
| Error (w) | 56 | 1,354.57 | | 55 | 1,220.64 | ($r_w$ = .34) |

Note.—St = modality of stimulus tested; T = test session (1 or 2).
\* $P < .05$.
\*\* $P < .01$.
\*\*\* $P < .001$.

$P < .001$, analysis of covariance). Hence, we were justified in making within experimental group and stimulus position comparisons.

$S_2$: *Group B vs. Group A.*—On the basis of our first hypothesis, we expected that Group B $Ss$, for whom $S_2$ was informative, should press more for $S_2$ than Group A $Ss$, for whom $S_2$ was redundant. The difference between the group means on the secondary reinforcing measure was in the predicted direction and significant beyond the .05 level ($F = 4.03$; $df = 1/56$). (The means are given in Table 1.) However, the effect was not statistically reliable in an analysis of covariance.

As mentioned above, 16 $Ss$, 8 each in Groups A and B, were trained with the number of occurrences of $S_1$ alone for Group B increased so that 80% of the stimulus events for Group B

were unaccompanied occurrences of $S_1$. For these $Ss$, tested with $S_2$ in Test Session 1, the means on the secondary reinforcing measure were in the predicted direction, 97.5 vs. 88.0, but the difference was short of statistical significance. However, when these data were analyzed in an analysis of covariance and combined by means of a critical ratio test with the data discussed above, the predicted effect was significant beyond the .05 level. ($CR = 1.97$ if the data from these 16 $Ss$ are combined with those from the 64 $Ss$ tested with $S_2$ in Test Session 1 or Test Session 2; $CR = 2.02$ if the data are combined with those from the 32 $Ss$ tested with $S_2$ in Test Session 1 only.)

$S_1$: *Group A vs. Group B.*—Our second hypothesis predicted that $S_1$ would be a more effective secondary reinforcer for Group A, for whom it

was reliable and informative, than for Group B, for whom it was unreliable. This prediction was borne out by the data beyond the .001 level ($F = 12.71$; $df = 1/55$; analysis of covariance).

*Group A: $S_1$ vs. $S_2$.*—As predicted from our first hypothesis, $S_1$ was a much more effective secondary reinforcer than $S_2$ for Group A. The difference between the means for these two stimulus positions, 115.1 vs. 65.8, was significant beyond the .001 level ($F = 26.35$; $df = 1/27$; analysis of covariance).

## CONTROL EXPERIMENTS

*Pseudoconditioned and unconditioned control.*—Fourteen $S$s, male albino rats, handled exactly as in the Main Experiment, were trained in groups of 7 $S$s each with stimulus sequences identical to those of Groups A and B, except that the stimuli were *never* paired with the occurrence of food, which was delivered at least 10 sec. after the occurrence of the stimuli. The two different patterns of stimuli used in training had no effect upon the pseudoconditioned rate of bar pressing. The mean for the 14 $S$s with both test sessions combined was 64.3. These 14 $S$s bar pressed for the stimuli significantly less in both Test Session 1 ($t = 3.41$; $df = 28$; $P < .005$) and Test Session 2 ($t = 2.72$; $df = 28$; $P < .02$) than did the 16 Group A $S$s bar pressing for the informative stimulus ($S_1$) in each of the Main Experiment test sessions. Hence, in a group predicted to show a large secondary reinforcing effect, we did indeed find such an effect produced by our training procedure.

Eight $S$s were exposed to the stimuli during training exactly as described above, except that the food pellets were eliminated entirely. The unconditioned rate of pressing for the stimuli was comparable to that of the pseudoconditioned group ($M = 73.4$).

The mean for the total group of pseudoconditioned and unconditioned $S$s with both test sessions combined was 67.6, indicating that the secondary reinforcing value of the redundant stimulus for Group A of the Main Experiment ($M = 65.8$), once the unconditioned rate of pressing for stimuli is taken into account, was small, if not zero, as we predicted from our first hypothesis. The estimates of the pseudoconditioned and un-

conditioned scores may be somewhat high, however, since these $S$s tended to have higher 10-min. extinction scores than did $S$s of the Main Experiment.

*Activation control.*—To test whether the effects studied in the Main Experiment were related to secondary reinforcement or only to a possible activation effect of a stimulus formerly associated with food (Wyckoff, Sidowski, & Chambliss, 1958), 10 additional $S$s were trained exactly as in the Main Experiment, 5 as in Group A and 5 as in Group B. However, during the testing of these $S$s, the bar remained nonfunctional once it was disconnected from the feeder. Each $S$ was tested at the same time as an identically trained $S$ used in the Main Experiment. The yoked Activation Control $S$ received only the stimuli earned by his Main Experiment partner. If the Main Experiment $S$ pressed for a stimulus within $7\frac{1}{2}$ sec. of a yoked Activation Control $S$'s response, the stimulus for the Activation Control $S$ was delayed so that it was not delivered until $7\frac{1}{2}$ sec. after his response. Hence spurious pairings of stimuli and pressing could not occur.

Thus, for these 10 $S$s, any pressing which occurred during the retraining test period could have been due only to the activation effects of the stimuli plus remaining operant level; the possibility of secondary reinforcement was eliminated.

In Test Session 1, all 10 of the Activation Control $S$s pressed less than did their secondary-reinforced partners ($P < .002$, binomial test, two-tailed). In Test Session 2, 9 out of 10 pressed less than did their yoked partners ($P < .02$, binomial test, two-tailed). Hence, we are quite certain that in the Main Experiment we were indeed studying secondary reinforcement.

*Partial reinforcement effect control.*—In the Main Experiment we had found that in the presence of a reliable predictor ($S_2$), training with partial reinforcement of $S_1$ produced less total pressing for $S_1$ as a secondary reinforcer than did 100% reinforcement. This confirmed our hypothesis but was opposite to the effect of increased resistance to extinction usually found with partial reinforcement. In order to see whether the presence of the reliable predictor was indeed the crucial factor, we ran two special control groups of 8 $S$s each, one with the usual partial reinforcement procedure and one with 100% reinforcement. These groups were identical in all respects to those of the Main Experiment, except that the reliable predictor, $S_2$, was omitted. When these groups were tested,

the partial reinforcement group tended to press more for the stimuli than did the continuous reinforcement group (though the difference between the group means, 128.6 vs. 115.6, was not statistically significant). However, the difference between these two groups was in the opposite direction and significantly different ($F = 5.71$; $df = 1/35$; $P < .025$) from the difference found between Test Session 1 means of the 32 Ss of the Main Experiment tested with $S_1$ during Test Session 1. Thus it appears that the presence of $S_2$, the reliable predictor of food, did play the crucial role in determining the direction of the results obtained in our tests of the secondary reinforcing value of $S_1$.

## DISCUSSION

Our situation differed from those in which the effect of partial reinforcement on the establishment of secondary reinforcement has been studied (e.g., Klein, 1959; Zimmerman, 1957, 1959) in that during training all our Ss had a reliable predictor of food. The seemingly crucial importance of the presence or absence of a reliable predictor during training may help to explain the apparently conflicting results obtained from single-group vs. separate-group experimental designs in determining the effects of partial reinforcement on the strength of a secondary reinforcer (e.g., D'Amato, Lachman, & Kivy, 1958). It may be that partial reinforcement will increase resistance to extinction of a secondary reinforcer only if training occurs in the absence of a reliable predictor.

It should be noted that our formulation of the conditions necessary for the establishment of a secondary reinforcer is compatible with the well-known "discriminative stimulus hypothesis" of secondary reinforcement (Keller & Schoenfeld, 1950; Schoenfeld, Antonitis, & Bersh, 1950). Furthermore, our results with respect to $S_2$: Group B vs. Group A could perhaps be considered analogous to those reported by Notterman (1951) in studies using rats as Ss in both a Skinner box and a straight alley.

## SUMMARY

Albino rats ($N = 88$, male) were trained to press a bar for food, then divided randomly into two groups and trained as follows for 135 trials in the same Skinner boxes with the bars removed: two stimuli, when paired, ended together and always preceded food. For Group A, the second, shorter stimulus ($S_2$) was always redundant because the first stimulus ($S_1$) had already given reliable information that food was to come. But for Group, B, $S_2$ was informative, because for them $S_1$ also occurred sometimes alone without food.

After the training sessions, the bars were reinserted, bar pressing was retrained with food pellets, extinguished, and then retrained again, this time using 1 sec. of one of the training stimuli as a secondary reinforcer in place of the food. The total number of bar presses in 10 min. following the first occurrence of the secondary reinforcing stimulus was used as the measure of secondary reinforcing strength. The testing procedure was repeated after 48 hr. using the other training stimulus as secondary reinforcer, so that all Ss were tested with both stimuli in a balanced sequence.

Control experiments were run to provide baseline levels for pseudoconditioned and unconditioned rates of pressing, and for any activating effect of the stimuli.

As predicted, $S_2$ was a stronger secondary reinforcer when it was informative than when it was redundant; $S_1$ was a more effective secondary reinforcer than $S_2$ in that group for which $S_2$ was a redundant predictor of primary reinforcement. In addition, $S_1$ was a more effective secondary reinforcer when it had been a reliable predictor of food.

## REFERENCES

D'AMATO, M. R., LACHMAN, R., & KIVY, P. Secondary reinforcement as affected by reward schedule and the testing situation. J. comp. physiol. Psychol., 1958, 51, 737–741.

HULL, C. L. Principles of behavior. New York: Appleton-Century, 1943.

KELLER, F. S., & SCHOENFELD, W. N. Principles of psychology. New York: Appleton-Century-Crofts, 1950.

KLEIN, R. M. Intermittent primary reinforcement as a parameter of secondary reinforcement. J. exp. Psychol., 1959, 58, 423–427.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951. Pp. 435–472.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior,

motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

MILLER, N. E. Analytical studies of drive and reward. *Amer. Psychologist,* 1961, **16,** 739–754.

NOTTERMAN, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *J. exp. Psychol.,* 1951, **41,** 161–169.

SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.,* 1950, **40,** 40–45.

SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

WYCKOFF, L. B., SIDOWSKI, J., & CHAMBLISS, D. J. An experimental study of the relationship between secondary reinforcing and cue effects of a stimulus. *J. comp. physiol. Psychol.,* 1958, **51,** 103–109.

ZIMMERMAN, D. W. Durable secondary reinforcement: Method and theory. *Psychol. Rev.,* 1957, **64,** 373–383.

ZIMMERMAN, D. W. Sustained performance in rats based on secondary reinforcement. *J. comp. physiol. Psychol.,* 1959, **52,** 353–358.

# INTERACTIONS AMONG THE SOMESTHETIC SENSES IN JUDGMENTS OF SUBJECTIVE MAGNITUDE [1]

## F. NOWELL JONES, DAVID SINGER, AND PAUL A. TWELKER

### University of California, Los Angeles

That perceptual blends occur among touch, warmth, and cold is well known (e.g., Boring, 1942). Interest in these perceptual phenomena has been enhanced recently by the revival of the suggestion that somesthetic senses of the skin are in fact not separate, but depend upon the same receptor and afferent mechanisms (Sinclair, 1955; Weddell, Palmer, & Pallie, 1955). Neurophysiological investigations support the view that some overlapping of mechanisms does occur, especially between touch and cold (Hensel & Boman, 1960). It is also true, for example, that the absolute thresholds for vibration may be altered by cooling or warming the skin (Weitz, 1941), although the physiological basis of the alterations is not clear.

Whether or not one can make judgments about one kind of stimulus dimension uncontaminated by the presence of a different kind of stimulus is another question. That the absolute threshold for a given modality can be changed by altering the condition of the skin or by improved coupling of the stimulator to the skin is not the case in point. Prolonged cooling or warming of the skin may change the threshold to touch by altering either the sensitivity of the receptors or by changing the biophysical characteristics of the skin. Changing the pressure of an applied cold or warm stimulator may alter the temperature thresholds by changing the conditions of coupling to the skin. Our concern is rather with the question of *S*'s ability to make judgments on the basis of information resulting from one variety of adequate stimulation in the presence of adequate stimulation of another kind. If he were able to make independent judgments, it would lend some support to the position of those who regard touch, cold, and warmth to be essentially independent channels of information.

## METHOD

*Apparatus.*—Since we proposed to present mechanical stimuli simultaneously with either cold or warmth, it was necessary to devise an apparatus which permitted controlled variation of these using the same stimulator for all. For this reason the apparatus finally devised used a temperature stimulator which could be either cooled or warmed, and which was attached to a device providing mechanical displacement for touch stimulation. The temperature stimulator is described in detail elsewhere (Jones, Twelker, & Singer, 1962), and consisted of a semiconductor thermocouple junction which could be cooled or warmed by the application of a direct current of proper orientation. The time constant, approximately 2 sec., was satisfactory for our purposes, as was the visual control of stimulus temperature by means of an imbedded thermocouple. The effective junction consisted of a copper disc 7.1 mm. in diameter which was used as the stimulus tip throughout the experiment. With this rather large stimulator it was not necessary to search for "temperature spots" to insure temperature stimulation. Mechanical displacement was achieved by means of a specially wound 500-ohm loudspeaker motor[2] whose move-

[2] The loudspeaker motor was made by Stephens Tru-Sonic, Incorporated, Culver City, California.

ment was controlled by an imposed current. Precise control of the amount, rate, and form of displacement was provided by feedback from the stimulator itself. Observed on an oscilloscope, the displacement of the stimulator was a straight line, with very sharp bends at the beginning and end of movement. There was no evidence of overshoot or ringing at the termination of displacement.

*Procedure.*—There were four experimental conditions: in Cond. P-W pressure was judged with concomitant warmth; in Cond. P-C pressure was judged with concomitant cold; in Cond. W-P warmth was judged with concomitant pressure; in Cond. C-P cold was judged with concomitant pressure. In each case, the primary stimulus dimension had five levels, the concomitant dimension had four. Pressure was defined as extent of stimulator movement, or depth of intrusion, this being the stimulus parameter previously selected as most useful (Jones, 1960). Temperature stimuli were defined as above or below skin temperature, which, needless to say, made it necessary to measure the skin temperature for every $S$ and adjust the stimulus series.[3] The $S$s differed from each other by as much as 2° C. For Cond. P-W and P-C the five pressure levels were 1, 2, 3, 4, and 5 mm. of rectilinear stimulus movement at a rate of 2 mm. sec. The four concomitant temperature stimuli were 0, 2, 5, and 9° C. above or below the measured skin temperature, respectively. For Cond. W-P and C-P, there were five temperature stimuli either above or below skin temperature by 0, 1, 3, 6, and 9° C. with four equally spaced concomitant pressure stimuli from 1 to 5 mm.

For any of the pairs of stimulus dimension there were 5 × 4 possible combinations of stimulus conditions. Since ordinal effects were expected, a 20 × 20 analysis of variance table was constructed according to the suggestions of Williams (1949). The design permitted the slightly confounded appraisal of first-order ordinal effects. Each row of the table represented 1 $S$, so that there was a total of 20 $S$s. The same $S$s served under all four conditions, but at four separate sittings. The orders of conditions were arranged as follows: for $S$s 1, 5, etc., P-C, C-P, W-P, P-W; for $S$s 2, 6, etc., W-P, P-W, P-C, C-P; for $S$s 3, 7, etc., P-W, W-P, C-P, P-C; for $S$s 4, 8, etc., C-P, P-C, P-W, W-P.

[3] Since heat transfer is a linear function of the absolute temperature differential, the actual skin temperature must be taken into account and is the logical zero point on the stimulus scale.

The stimuli were judged by the method of magnitude estimation (Stevens, 1957), since this had proved to be a useful method in previous work on pressure (Jones, 1960). In Cond. P-W and P-C, $S$s were instructed to judge on the basis of pressure or touch, while in Cond. C-P and W-P they were instructed to judge on the basis of coldness and warmth, respectively. For pressure, a standard stimulus movement of 3 mm. was designated as 10. In the case of a temperature a 3° C. deviation from skin temperature with a stimulus movement of 1 mm. was designated as 10. In every case the standard was presented three times at the beginning of the experimental session, and not afterwards repeated.

Three special problems of stimulus control require comment. First, with temperature stimulation there is a tendency for the skin temperature to drift as the result of stimulation. This was avoided by returning the stimulator to the previously measured skin temperature after each stimulation. Second, the timing of combined pressure and temperature stimuli is complicated by the shorter time constant of the skin for mechanical as compared to thermal stimuli. In the present experiment, the concomitant temperature change was initiated approximately 5 sec. before the mechanical stimulus. Third, the ambient temperature remained within 1° F. of 72° F. throughout the experiment.

*Subjects.*—The $S$s were drawn from one beginning and one advanced class in psychology. They were not experienced in psychophysical judgment, but very few had any difficulty. A few $S$s were eliminated because of failure to understand the instructions or because of difficulties with the apparatus.

## RESULTS

After a logarithmic transformation, the data for each of the four conditions were analyzed for ordinal effects according to the routine given by Cochrane and Cox (1957, pp. 135–138). There was no significant first-order effect for any condition, hence no correction for order was made. Tables 1 and 2 give the analyses of variance for the four conditions. As might be expected (Jones et al., 1961), there is a highly significant $S$ effect, using a conservative error term

(triple interaction), in every condition, and the $S \times$ Primary Modality interactions are significant, indicating differences in slope. The primary effect of the dimension to be judged is highly significant in every case, again as might be expected. In no case is the effect of the concomitant stimulus modality significant. In only one case, Cond. C-P, is the interaction between modalities significant. Inspection of the Cold $\times$ Pressure table indicates that this interaction results from a slight tendency for the intermediate but not extreme cold stimuli to be judged greater when accompanied by a greater pressure. An analysis was performed to discover whether or not there was an overall effect of cold or warmth on judgments of pressure, since this could not be determined from the analysis of each condition separately. The result shows that there is no significant effect, pressures being judged the same whether accompanied by cold or warmth ($F = .10$, $df = 1/19$).

## DISCUSSION

That the somesthetic senses somehow interact to provide perceptual patterns

#### TABLE 1

ANALYSIS OF VARIANCE OF MAGNITUDE ESTIMATES OF PRESSURE MADE WITH CONCOMITANT WARMTH OR COLD

| Source | df | Concomitant Warmth (W) | | Concomitant Cold (C) | |
|---|---|---|---|---|---|
| | | MS | F | MS | F |
| Ss | 19 | .48 | 24.0** | .59 | 19.3** |
| Warmth (or Cold) | 3 | .02 | .66 | .04 | 1.3 |
| Pressure (P) | 4 | 3.63 | 45.5** | 4.56 | 91.2** |
| Ss × W (or C) | 57 | .03 | 1.5* | .03 | 1.0 |
| Ss × P | 76 | .08 | 4.0** | .05 | 1.6* |
| P × W (or C) | 12 | .02 | 1.0 | .04 | 1.3 |
| P × W (or C) × Ss | 228 | .02 | | .03 | |

\* P = .05.
\*\* P = .01.

#### TABLE 2

ANALYSIS OF VARIANCE OF MAGNITUDE ESTIMATES OF WARMTH AND COLD MADE WITH CONCOMITANT PRESSURE

| Source | df | Estimates of Warmth | | Estimates of Cold | |
|---|---|---|---|---|---|
| | | MS | F | MS | F |
| Ss | 19 | 1.02 | 25.5** | .83 | 41.5** |
| Pressure (P) | 3 | .05 | 1.2 | .10 | .71 |
| Warmth (W) or Cold (C) | 4 | 3.95 | 23.2** | 6.38 | 35.4** |
| Ss × P | 57 | .04 | 1.0 | .14 | 7.0** |
| Ss × W (or C) | 76 | .17 | 4.2** | .18 | 9.0** |
| W (or C) × P | 12 | .05 | 1.2 | .05 | 2.5** |
| Ss × W (or C) × P | 228 | .04 | | .02 | |

\*\* P ≤ .01.

such as "wet," "oily," and the like is a commonplace observation which was at one time, at least, the subject of careful study (cf. references in Boring, 1942, pp. 521–522). Also indicative of interaction is the tendency for cold objects to be judged heavier than warm ones (Weber, 1905). From neurophysiological work, including the recording of action potentials from human nerve fibers (Hensel & Boman, 1960), we know that cold stimulation and pressure stimulation affect, in part, the same fibers, and we also know that there is some overlapping of representation in the cerebral cortex (Landgren, 1957). Even without accepting in detail the arguments of Weddell et al. (1955) and Sinclair (1955), among others, that there is no real differentiation of the skin senses, we should still expect, on the basis of both the perceptual and neurophysiological literature, some significant degree of interaction. We have found very little, if any. Pressure can be judged independently of concomitant temperature stimulation, and cold and warmth can be judged independently of concomitant pressure stimulation, with the possible exception of a small interaction of cold and pressure.

Psychologically, that is, as far as the total system response is concerned, there is very little to say about the results beyond their demonstration and the suggestion that further work may refine the results presented here, es-

pecially in regard to the effects of changing the temporal relationships among the stimuli. The neurophysiological implications, however, require comment. In the first place, the possibility of independence of magnitude judgments requires that there be differentiation among receptors. As Brindley (1957) has pointed out for the visual system, the nervous system cannot create information out of an undifferentiated peripheral response, no matter how complex the subsequent transformations. Support is lent, therefore, to the position that separate receptors are involved in the response to different varieties of stimulation. In the second place, it appears that the judgments of magnitude are made on the basis of some steady state in the case of temperature, rather than upon the initial burst of impulses obtainable not only from cold receptors but from presumed pressure receptors associated with large axones upon the application of cold (Hensel & Zotterman, 1951). It is possible that different temporal relationships among the stimuli would lead to the discovery of interactions that our particular choice of timing has not revealed, but for the present such a suggestion is purely speculative. A third, and final point is that the overlapping in cortical representation of the various skin modalities, even though involving convergence on the same cortical neuron in some instances, somehow does not interfere with modality-specific judgments. Furthermore, the facilitatory effect of pressure stimulation suggested by Landgren (1957) has no psychological counterpart in our results. A consideration of the latencies involved would tentatively suggest that the nonspecific cells found in the cortex are related to the arousal system rather than to a specific sensory pathway, and would not be expected to be directly involved in judgments of magnitude.

We may conclude, therefore, that our results are compatible with the existence of separable mechanisms of response to pressure and thermal stimulation. Although we cannot argue for the accept-ance of any particular neurological mechanism as underlying our results, the ideas advanced here are parsimonious and in harmony with the weight of neurophysiological evidence. In any case, no matter what the underlying mechanism, we have shown that pressure, cold, and warmth can be responded to selectively in the presence of concomitant stimulation.

## SUMMARY

Twenty Ss gave magnitude estimates of pressure stimuli in the presence of concomitant cold or warm stimuli, and magnitude estimates of cold and of warmth in the presence of concomitant pressure stimuli. It was found that judgments of magnitude can be made independently of concomitant stimulation in another modality. It was suggested that this result is consistent with the assumption of separable neurological mechanisms for the skin senses under consideration.

## REFERENCES

BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.

BRINDLEY, G. S. Two theorems in colour vision. *Quart. J. exp. Psychol.,* 1957, **9**, 101–104.

COCHRAN, W. G., & COX, G. M. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

HENSEL, H., & BOMAN, K. K. A. Afferent impulses in cutaneous sensory nerves in human subjects. *J. Neurophysiol.,* 1960, **23**, 564–578.

HENSEL, H., & ZOTTERMAN, Y. The response of mechanoreceptors to thermal stimulation. *J. Physiol.,* 1951, **115**, 16–24.

JONES, F. N. Some subjective magnitude functions for touch. In G. R. Hawkes (Ed.), *Symposium on cutaneous sensitivity.* Ft. Knox, Ky: United States Army Medical Research Laboratory, 1960. (No. 424)

JONES, F. N., TWELKER, P. A., & SINGER, D. A waterless thermal somesthetic stimulator. *Amer. J. Psychol.,* 1962, **75**, 147–149.

LANDGREN, S. Convergence of tactile, thermal, and gustatory impulses on single cortical cells. *Acta physiol. Scand., Stockh.,* 1957, **40**, 210–221.

SINCLAIR, D. C. Cutaneous sensation and the doctrine of specific nerve energy. *Brain*, 1955, **78**, 584–614.

STEVENS, S. S. On the psychophysical law. *Psychol Rev.*, 1957, **64**, 153–181.

WEBER, E. H. *Der Tastsinn und das Gemeingefühl*. Leipzig: Englemann, 1905.

WEDDELL, G., PALMER, E., & PALLIE, W. Nerve endings in mammalian skin. *Biol. Rev.*, 1955, **30**, 159–195.

WEITZ, J. Vibratory sensitivity as a function of skin temperature. *J. exp. Psychol.*, 1941, **28**, 21–36.

WILLIAMS, E. J. Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. scient. Res. Ser. A*, 1949, **2**, 149–168.

# POSTCONDITIONING DELAY AND INTENSITY OF SHOCK AS FACTORS IN THE MEASUREMENT OF ACQUIRED FEAR [1]

WALLACE R. McALLISTER AND DOROTHY E. McALLISTER

*Syracuse University*

That the presentation of a stimulus previously paired with electric shock can serve as the basis for the learning of another response is a well-established finding. The most convincing evidence is provided by a two-stage experiment. First, a neutral CS is paired with inescapable electric shock (UCS). Then the formerly neutral stimulus is presented without the shock, and Ss are allowed to make another (indicator) response which leads to the termination of the CS. Examples of such experiments are those of Brown and Jacobs (1949) and Kalish (1954). A variation of these procedures was used in the original study of Miller (1948). Theoretically, it is assumed that fear, as a response, is classically conditioned to the neutral CS in the first stage of the experiment. In the second stage, fear elicited by the CS serves as a motivator and its decrease, occurring when the CS is terminated, acts as a reinforcer for the learning of the indicator response.

Despite the number of experiments which have supported the theory, difficulty in obtaining evidence for the acquired drive of fear with the above procedures has been reported (Brown & Jacobs, 1949; Solomon &

Brush, 1956, p. 221). Two preliminary experiments conducted in this laboratory, in which hurdle jumping was used as the indicator response, likewise yielded negative results. As a first step in attempting to account for these results, a study designed to determine the optimal shock intensity for conditioning was begun. The initial results indicated that, regardless of the shock level employed, the hurdle-jumping response was not learned. Subsequent investigation revealed, however, that learning became evident when additional trials were given on the following day. Therefore, the study was redesigned and conducted as reported in Exp. I. The purpose of Exp. II was to explicate the finding of Exp. I that learning occurred only on the second day of training.

## EXPERIMENT I

The purpose of this study was to investigate the effect on hurdle-jumping performance of the intensity of shock used during fear conditioning.

### Method

*Subjects.*—The Ss were 100 naive, female, hooded rats from the colony maintained by the Psychology Department at Syracuse University. Nine additional Ss were discarded, 8 for apparatus failure and 1 because of extreme difficulty in handling during hurdle jumping. The Ss were randomly paired and five pairs were then assigned at random to each of 10 groups. Their ages ranged from 103 to 258 days at the start of the experiment. The distribution of ages was approximately the same for the groups, with means varying between 143 and 168 days.

*Apparatus.*—Two shock boxes and a hurdle-jumping apparatus were used. The hurdle-jumping apparatus consisted of two boxes, $9\frac{1}{4}$ in. long $\times$ $4\frac{1}{2}$ in. wide $\times$ 5 in. high (interior dimensions), separated by a $\frac{3}{4}$-in. partition containing a guillotine door, $2\frac{1}{2}$ in. wide $\times$ 3 in. high, which rested on a hurdle 2 in. high. The grid box, painted white, had a floor made of $\frac{3}{32}$-in. brass welding rods spaced $\frac{7}{16}$ in. apart. The safe box, painted gray, had a plywood floor constructed so as to serve as a floor switch. A .01-sec. Standard Electric timer was started with the opening of the guillotine door and was stopped by depression of the floor.

The grid and safe boxes each had another box, $13\frac{1}{4}$ in. high, hinged to its top which contained light sources and acted as a cover. The bottom of each upper box was covered with hardware cloth, 1 in. above which was inserted a pane of opal glass. A $7\frac{1}{2}$-w. lamp located in each of the upper boxes and a 40-w. lamp located in the box above the grid provided the intertrial and CS illuminations.

So that Ss could be run in pairs, two separate shock boxes, wired in parallel, were used for the fear-conditioning phase of the experiment. They were constructed to appear identical with the grid box of the hurdle-jumping apparatus. Hunter interval timers were used to control the presentations of the CS and the UCS; a Haydon timer controlled the duration of the intertrial interval. Except for the use of a variac, monitored by a voltmeter, to control the level of shock to the grids, the circuit employed was that described by Wyckoff and Page (1954). A 100,000-ohm resistor was in series with each S. The grids were energized successively at a rate of two impulses, of about 13-msec. duration each, per grid per sec.

*Design and procedure.*—The procedures for each S required 4 days. The first 2 days were devoted to handling and to exploration of the hurdle-jumping apparatus. On each of these days each S was handled for two 10-min. sessions and was allowed to explore each side of the hurdle-jumping apparatus for 10 min. with the guillotine door closed. During handling, E was seated in front of a table, $34\frac{3}{4} \times 17\frac{3}{4}$ in., enclosed on three sides with curtains. Handling consisted of alternately picking up and stroking S and placing her on the table to explore. The sequence of treatment for 1 S of the pair was handling, exploration of the grid box, handling, and exploration of the safe box, for 10 min. each. Since Ss in a pair were treated concurrently, the sequence for the other member began with exploration of the grid box. During the first 5 min. of handling on the first day and the last 5 min. on the second day, S remained on the table and was not touched so that an activity measure could be obtained. These data will not be discussed in this paper.

On the third day, Ss in each of five groups were given 35 forward-conditioning (FC) trials. The CS was an increase in illumination of 6-sec. duration from 7 to 115 ft-c, measured with a Weston illumination meter, Model 756. The UCS was a shock of 2-sec. duration which was presented 4 sec. after the onset of the CS. The intensity of shock delivered to the groups was either 30, 40, 50, 60, or 100 v. Five other groups were given 35 backward-conditioning (BC) trials, one group at each shock level. For these groups 15 sec. intervened between the offset of the shock and the onset of the CS. For BC Ss, trials were started 10 sec. after placement in the shock boxes. To equalize the amount of time spent in them, FC Ss were placed in the boxes 605 sec. before their first trial. A 2-min. intertrial interval was used for all groups.

Ten seconds after the last conditioning trial, Ss were removed to separate holding boxes. Approximately 2 min. later 25 hurdle-jumping trials were begun. On each trial S was placed in the grid box facing the guillotine door which was raised after 10 sec. The CS was presented simultaneously with the raising of the door and was terminated by depression of the floor switch when S crossed the hurdle into the safe box. No shock was administered during this phase of the experiment. When S jumped, the door was closed, and after 10 sec. S was returned to the holding box. If no jump occurred within 60 sec., S was removed to the holding box and a latency of 60 sec. recorded. The Ss were run alternately with a minimum intertrial interval of 30 sec. After the trials, Ss were weighed.

On the fourth day, 25 additional hurdle-jumping trials were given. On this day, if S failed to jump within 60 sec. on 10 consecutive trials, training was terminated and 60 sec. was recorded for each of the remaining trials. Throughout this paper, the 2 hurdle-jumping days will be referred to as Day 1 and Day 2.

Food and water were available at all times in the home cages.

## Results

In Fig. 1 the means of the reciprocals of latency of hurdle jumping in five-trial blocks are plotted for each group. No evidence of learning ap-
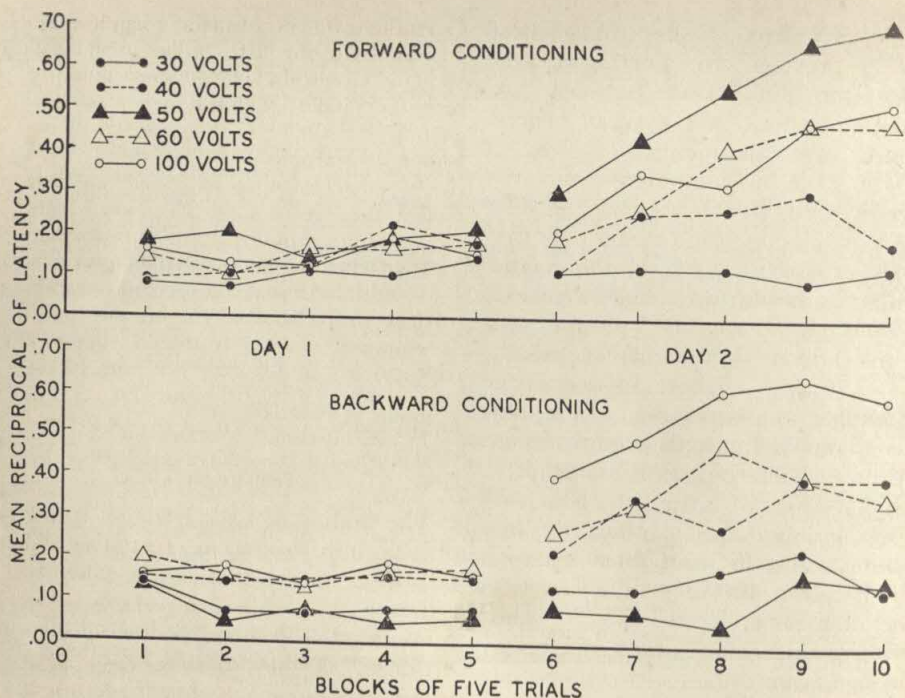
FIG. 1. Mean reciprocal of latency of hurdle jumping as a function of blocks of five trials following forward and backward conditioning at each shock level.

pears until Day 2 of hurdle jumping when, in general, the performances of those groups conditioned with the higher levels of shock improve. Except for the unusual performances of both 50-v. groups, for which no satisfactory explanation can be offered, there is an increasing monotonic relationship between mean performance on Day 2 and intensity of shock for both FC and BC groups. The BC groups (except the one with 50 v.) performed as well as their FC counterparts.

Because the assumption of homogeneity of variance was untenable, a trend analysis of variance was not used. To evaluate the changes in performance over the hurdle-jumping trials, a $t$ test for related measures was computed for each group using the means of reciprocals of latency

on Trial Blocks 1 and 10. Using a 5% coefficient of risk, adopted for all statistical tests reported in this paper, the 50- and 60-v. FC groups showed a significant gain in performance, while that for the 100-v. FC group approached significance ($t = 2.86$, $3.18$, and $1.93$, $df = 9$, respectively). Of the BC groups, only the 100-v. group showed a significant gain ($t = 2.80$, $df = 9$).

To determine the effects of intensity of shock and type of conditioning, the performance measures on Trial Block 10 were transformed, because of the failure to meet the assumption of homogeneity of variance, and then subjected to a factorial analysis of variance. The measures for this block of trials were ranked over all groups and then converted to mean deviations using an extension of

Table XX from Fisher and Yates (1957) provided by Porter (1958). Since the interaction between intensity of shock and type of conditioning was significant ($F = 3.31$, $df = 4/90$), simple analyses of variance were used to evaluate the differences among the means of the shock groups for each of the conditioning procedures. For the FC groups and for the BC groups, the means differed significantly ($F = 5.22$ and $2.76$, $df = 4/45$, respectively). A comparison was then made between the individual group means within each of the conditioning procedures. For FC, the 50-, 60-, and 100-v. groups were significantly superior to the 30- and 40-v. groups. For BC, the 100-v. group was significantly superior to the 30- and 50-v. groups. No other difference was significant.

A further analysis was made between FC and BC groups for each level of shock. The 50-v. FC group was significantly superior to the 50-v. BC group. Since none of the other differences was significant and since both 50-v. groups performed in an unusual manner, it is difficult to interpret this finding.

## Discussion

The general tendency for learning to improve with increases in the intensity of shock used during fear conditioning is in agreement with previous findings (Goldstein, 1960; Miller, 1951). Such a result would be expected since, presumably, the amount of fear conditioned is related to the level of shock which, thereby, would determine the degree of learning of the indicator response.

Two results are not consistent with those usually obtained. (a) Learning of the indicator response was not evidenced until late in training (on Day 2). This finding will be discussed in connection with the results of Exp. II. (b)

Learning of the indicator response occurred following BC procedures. Although Goldstein (1960) has recently reported similar findings, other investigators (Kalish, 1954; Porter, 1958) have found no learning in BC groups. Most likely such learning is based on fear conditioned to the apparatus cues which, when strong shock is used, is not extinguished during the intertrial intervals. It should be noted that a period permitting extinction of fear to the apparatus cues was introduced between conditioning and hurdle jumping in the experiments of Kalish and Porter but not in that of Goldstein or in the present study.

## EXPERIMENT II

The finding of Exp. I that learning of the hurdle-jumping response did not occur until Day 2 can be taken to indicate that either the elapse of a minimal time interval following conditioning or the administration of a minimal number of hurdle-jumping trials is a necessary condition for learning. The present experiment was designed to allow a choice between these two alternatives. For some Ss there was a 1-day delay between conditioning and hurdle jumping while for other Ss there was no delay. If the number of trials is the important variable, no difference would be expected between the delay and no-delay conditions. If, on the other hand, the postconditioning delay interval is crucial, learning should occur on the first day of hurdle jumping following the delay.

## Method

*Subjects.*—The Ss were 40 naive, female, hooded rats ranging between 98 and 111 days of age at the beginning of the experiment. The source of Ss and the method of pairing and assigning Ss to groups were the same as in Exp. I.

*Apparatus.*—The same apparatus was used as in the previous experiment.

*Design and procedure.*—Four groups of 10 Ss each, two FC and two BC, were used. All

procedures were the same as in Exp. I except that the shock level was 70 v. for all groups and that for two groups (FC-D and BC-D) there was a delay of approximately 22½ hr. between the completion of conditioning and the beginning of hurdle jumping. For the other two groups (FC-ND and BC-ND), there was no delay beyond the minimal time needed to prepare the apparatus (approximately 2 min.).

*Results*

The means of the reciprocals of latency of hurdle jumping in blocks of five trials for each of the groups are presented in Fig. 2. The superiority of performance of the FC-D group as compared with the other three groups is clearly evident on Day 1 and is maintained on Day 2. The BC groups and the FC-ND group appear to perform similarly and at about the same level throughout training. For the statistical analysis of the differences between the groups, the reciprocals of latency on Trial Block 5 were transformed in the manner described in Exp. I and then subjected to a factorial analysis of variance with Delay (D or ND) and Type of Conditioning (FC or BC) as the factors. Since the interaction was significant ($F = 9.03, df = 1/36$), the simple effects were analyzed. For FC the D group performed better than the ND group ($t = 4.45, df = 36$) while for BC the groups did not differ. For the D condition, the FC group performed significantly better than the BC group ($t = 3.95, df = 36$) while for the ND condition the groups did not differ. Thus, under the conditions of this experiment, the Delay variable was effective only following FC while the Conditioning variable was effective only when a delay was used.
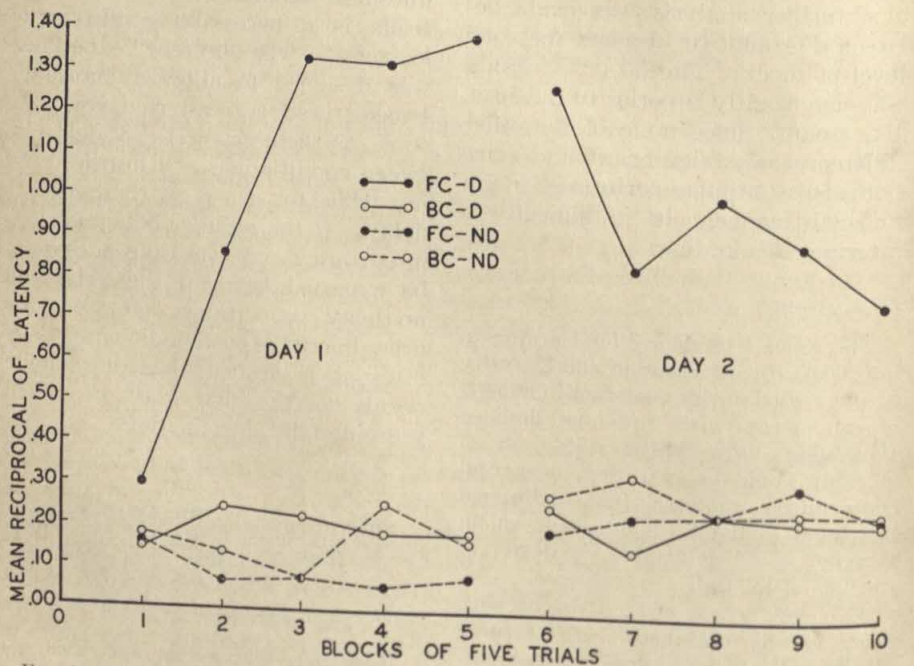


FIG. 2. Mean reciprocal of latency of hurdle jumping as a function of blocks of five trials following forward conditioning (FC) and backward conditioning (BC) with a delay (D) or with no delay (ND) between conditioning and hurdle jumping.

The changes in performance from Trial Blocks 1 to 5 were evaluated with $t$ tests for related measures. Only Group FC-D showed a significant improvement in performance ($t = 4.36$, $df = 9$). The failure to find any evidence for learning in Group BC-D probably can be attributed to the level of shock used since in later experiments, when higher levels of shock were employed, groups trained under that condition did learn. No evidence of learning was shown by the ND groups on either day. The lack of learning on Day 2 in Group FC-ND is inconsistent with the results of Exp. I in which Group FC with a comparable level of shock (60 v.) did learn. From Fig. 1 and 2, however, it can be seen that the performances of these two FC groups are similar through the seventh block of trials. Only on the last three blocks of trials is the performance of the Exp. I group noticeably higher. Since the absolute difference in performance is not great, it is probably most parsimonious to attribute the discrepancy to sampling error. The results for the BC groups with comparable shock levels from the two experiments are consistent since learning was not shown in either case.

## Discussion

In this experiment, learning of the hurdle-jumping response occurred early in training only with an adequate postconditioning delay. Thus, the failure to obtain any evidence of learning on Day 1 in Exp. I can be attributed to the use of an inadequate delay rather than to the lack of a sufficient number of hurdle-jumping trials. The results are consistent with those of other investigators (Brown & Jacobs, 1949; Kalish, 1954; Kent, Wagner, & Gannon, 1960; Porter, 1958) who have reported rapid learning of the hurdle-jumping response. A similar delay was employed in these studies following all, or the major portion of, the conditioning trials, but its use seems to have been fortuitous, and the relevance of this temporal variable in this type of experiment seems not to have been recognized. The existing data do not allow a determination of the critical length of delay.

No evidence is provided by the present data concerning the nature of the events occurring during the postconditioning delay which leads to its effect on hurdle-jumping performance. Several possible interpretations can, however, be mentioned. Two of these assume that the strength of the fear response varies in time. One example is the incubation of fear hypothesis (Bindra & Cameron, 1953, p. 197). Another, based on the findings of Perkins and Weyant (1958), assumes that the strength of the fear response to *generalized* stimuli increases during the postconditioning delay. Since conditioning and hurdle jumping were carried out in different, although similar, apparatuses, stimulus generalization might be an important factor. Other interpretations rely on the variation in time of the strength of some factor which interferes with the operation of the fear response. For instance, it may be assumed that following conditioning a strong, general emotional state, which dissipates with time, is present as an aftereffect of shock (Amsel & Maltzman, 1950). In this case, immediately following conditioning, hurdle jumping might not result in a sufficient decrease in total emotionality to be reinforced. On the other hand, it could be assumed that the cues resulting from this general emotionality elicit some response incompatible with hurdle jumping, such as crouching, which would interfere with the learning until the emotionality was dissipated. At this time there are no firm grounds for choosing between these alternatives.

## SUMMARY

Two experiments concerned with the classical conditioning of fear in rats and the measurement of its effect through the learning of another response (hurdle jumping)

were conducted. In Exp. I the effect of intensity of shock (30, 40, 50, 60, or 100 v.) used during conditioning was investigated. At each shock level one group was given forward conditioning (light-shock) and one, backward conditioning (shock-light). For all groups hurdle-jumping trials in which *S* could escape the light by jumping a hurdle were administered immediately following conditioning and were continued on the next day. Evidence of learning was obtained following both forward and backward conditioning but only on the second hurdle-jumping day. Performance, in general, was better following conditioning with the higher shock levels. The results of Exp. II indicated that learning does occur on the first day of hurdle jumping when a postconditioning delay of 1 day is used.

## REFERENCES

AMSEL, A., & MALTZMAN, I. The effect upon generalized drive strength of emotionality as inferred from the level of consummatory response. *J. exp. Psychol.*, 1950, **40**, 563–569.

BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with the passage of time: Incubation effect. *J. exp. Psychol.*, 1953, **45**, 197–203.

BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research.* London: Oliver & Boyd, 1957.

GOLDSTEIN, M. L. Acquired drive strength as a joint function of shock intensity and number of acquisition trials. *J. exp. Psychol.*, 1960, **60**, 349–358.

KALISH, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–9.

KENT, N. D., WAGNER, M. K., & GANNON, D. R. Effect of unconditioned response restriction on subsequent acquisition of a habit motivated by "fear." *Psychol. Rep.*, 1960, **6**, 335–338.

MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.

PERKINS, C. C., JR., & WEYANT, R. G. The interval between training and test trials as a determiner of the slope of generalization gradients. *J. comp. physiol. Psychol.*, 1958, **51**, 596–600.

PORTER, L. G. Generalization of fear and of the inhibition of fear. Unpublished doctoral dissertation, Syracuse University, 1958.

SOLOMON, R. L., & BRUSH, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 212–305.

WYCKOFF, L. B., & PAGE, H. A. A grid for administering shock. *Amer. J. Psychol.*, 1954, **67**, 154.

# OVERLEARNING AND POSITION REVERSAL [1]

## M. R. D'AMATO AND H. JAGODA

### *New York University*

In a previous paper (D'Amato & Jagoda, 1961) it was shown that the overlearning effect (the facilitation of discrimination reversal by extended postcriterion training) did not occur in a visual discrimination task if Ss were required (via forced trials) to have a moderate amount of experience with the negative stimulus (S_) during the postcriterion training. The interpretation of this result was that continued experience with S_ during overlearning prevented a reduction in avoidance of that stimulus (as normally occurs, presumably, during extended overtraining), and with avoidance of S_ maintained at a relatively high level, facilitation of reversal learning was precluded.

The purposes of the present studies were twofold. First, the preceding analysis of the overlearning effect is not as easily applied to position discrimination reversal, where there exist no positive and negative stimuli as such. And yet, the overlearning effect has been reported in a position discrimination task (Pubols, 1956). One aim of the present studies, therefore, was to determine whether, in position reversal, a moderate proportion of forced trials to the incorrect side during postcriterion training would similarly result in the disappearance of the overlearning effect. Second, in most overlearning studies (e.g., Birch, Ison, & Sperling, 1960; Bruner, Mandler, O'Dowd, & Wal-

lach, 1958; Pubols, 1956; Reid, 1953) the overtraining trials were distributed over an extended period, some 12 to 30 days, while the control Ss were reversed immediately upon reaching acquisition criterion. Since there is some evidence that mere delay between acquisition and reversal serves to facilitate reversal learning (Bunch, 1939; Murofushi, 1957; Stevenson & Weir, 1959), the evaluation of this variable by the addition of an appropriate control group would seem advisable; such a control group is incorporated in two of the studies reported below.

## EXPERIMENT I

### *Method*

*Subjects and apparatus.*—The Ss were 72 experimentally naive albino rats (25 males and 47 females), 80 to 130 days of age at the start of the study. Four automatic Y maze discrimination apparatuses, fully described elsewhere (D'Amato & Jagoda, 1960), were used.

*Pretraining.*—Four days prior to the beginning of acquisition, Ss were placed on a standard 23-hr. water deprivation regimen. Each S was placed in a darkened arm of the maze and permitted to drink five dipperfuls of water on the day preceding the beginning of acquisition; three dipperfuls were allowed on the subsequent day.

*Position acquisition.*—All Ss were trained on a right position response. Ten trials were given on Day 1, S being forced on Trial 2 to the side opposite to that chosen on Trial 1; on Trial 7, S was forced to the same side as that chosen on Trial 1. On all subsequent acquisition days, 20 trials per day were given, with 1 out of every 5 trials forced; 2 of the 4 daily forced trials were to the incorrect (left) side. In order to facilitate generalization between free and forced trials, transparent plastic doors were used.

On a correct choice the sequence of events was as follows. The S was permitted 1.75 sec.

±.25 sec. drinking (reward interval), timed from $S$'s first lap of water, after which the dipper retracted out of sight. The lights in the maze remained on for an additional 8 sec. A 45-sec. ± 3 sec. intertrial interval followed, during which all lights in the maze were extinguished. Finally, a flashing light (60 flashes per min.) of 4.5 sec. ± .5 sec. duration signaled the beginning of the next trial. On an incorrect choice the sequence of events was identical, except that the dipper was retracted during the reward interval. The illumination in the right and left arms was identical and, using our standard procedure, measured 1.0 ft-c. It should be noted that, unlike most choice situations, no extramaze cues are available in our apparatus; thus, the position response, being uncorrelated with any cue, is very nearly a "pure" turning response.

Fifteen to 30 min. after the day's trials, $S$s were watered for 1 hr. During acquisition, overlearning, and reversal the estimated average duration of water deprivation at the start of a day's trials was 22.5 hr.

As $S$s reached acquisition criterion (18 correct responses out of any 20 successive trials) in squads of 4, they were assigned at random to one of the following four groups (18 $S$s per group). Group C was the standard control group, $S$s of this group being placed on reversal the day after reaching criterion. Group CD was the added control group that had the beginning of its reversal training delayed the same length of time (8 days) occupied by the postcriterion training given to the overlearning groups. During these 8 days, Group CD was maintained on its regular 23-hr. water deprivation schedule. Groups EC and EI were experimental groups that received the overlearning experience described below.

*Overlearning and reversal.*—Groups EC and EI received 160 overlearning trials, 20 per day for 8 days, with 4 (20%) of each day's trials being forced. For Group EC, all forced trials were to the correct (right) side; for Group EI, all forced trials were to the incorrect (left) side. Apart from this difference, the two groups received identical treatment during overlearning.

For all groups reversal training consisted of 20 trials per day, all free. Reward, of course, now appeared in the left arm of the maze. The reversal criterion was the same as employed during acquisition.

## Results and Discussion

*Acquisition and overlearning.*—The mean numbers of free trials to acquisition criterion were 38.1, 40.3, 39.4, and 40.1, for Groups C, CD, EC, and EI, in that order. Group EC averaged 5.6% errors on the 128 free overlearning trials, while for Group EI the corresponding value was 9.9%, the difference between the groups being nonsignificant.

*Reversal.*—Using the individual (trials to criterion) acquisition scores, $S$s of each of the four groups were divided into three levels (6 $S$s per level), reflecting the speed with which the acquisition criterion was attained. A Treatments × Levels analysis of variance was then applied to the trials to reversal criterion data. The $F$ for treatments alone proved significant ($F = 21.12$; $df = 3/60$; $P < .01$). In the pairwise $t$ tests that were applied to the group means (which appear in the first row of Table 1), Group EI emerged significantly inferior to all other groups ($P < .001$ in all comparisons). However, all other pairwise differences fell far short of significance. Similar analyses based on the mean numbers of errors to reversal criterion yielded similar results.

In agreement with previously reported results (D'Amato & Jagoda, 1961), the data of the present experiment clearly show that forced trials to the incorrect side during overlearning sharply interfere with position reversal learning. However, it is equally clear that the overlearning effect has not been obtained.

One possible explanation for the absence of the overlearning effect is suggested by the results of Group CD, the group that had reversal training delayed for 8 days. It will be observed from Table 1 that this group had the smallest mean number of trials to reversal criterion, though not significantly so. Since, as earlier mentioned, many studies reporting the overlearning effect have distributed the overtraining trials over a minimum of 12 days, and since there is some evidence that delay between acquisition and reversal can facilitate reversal learning, the possibility arises that the overlearning effect is some joint

## TABLE 1

MEAN NUMBERS OF TRIALS TO REVERSAL
CRITERION

| Exp. | Group | | | | N per Group |
| --- | --- | --- | --- | --- | --- |
| | C | CD | EC | EI | |
| I | 76.4 | 66.6 | 85.8 | 181.9 | 18 |
| II | 91.2 | 79.4 | 93.9 | 177.6 | 10 |
| III | 86.0 | — | 101.8 | 157.0 | 13 |

function of number of overtraining trials and the time elapsed between achievement of acquisition and the beginning of reversal learning. There is a plausible similarity between the operation of overlearning and that of interposing a delay between acquisition and reversal learning. In the latter, S is removed from the experimental situation for a period of time and therefore has no experience during this delay with either the positive or the negative stimulus. In the former, since few errors are normally made during overtraining, experience with the negative stimulus (or with the incorrect response) can be considered to be virtually terminated during the greater part of the overlearning phase.

Thus, a second experiment was run which was closely similar to Exp. I with the important difference that 10 rather than 20 trials per day were given through all phases of the experiment. The 160 postcriterion trials therefore required 16 rather than 8 days.

## EXPERIMENT II

### Method

*Subjects and apparatus.*—Forty experimentally naive albino Ss (23 males and 17 females) 85 to 100 days of age were used. In this study opaque rather than transparent doors were used, a change necessitated by the requirements of other, concurrently run, studies.

*Procedure.*—The only differences in the procedure of this and the preceding experiment were that (a) 10 trials per day were given during acquisition, overtraining, and reversal; and (b) the intertrial interval was increased from 45 to 60 sec. Two (20%) of the daily acquisition and overlearning trials were forced.

The 40 Ss were assigned to Groups C, CD,

EC, and EI, 10 Ss per group, in such a manner as to equate for acquisition means.

### Results and Discussion

*Acquisition and overlearning.*—The mean numbers of free trials to acquisition criterion for Groups C, CD, EC, and EI, respectively, were 57.7, 53.6, 53.1, and 50.1. Of the 128 free postcriterion trials Group EC averaged 8.1% errors and Group EI, 4.8%. The difference between the group means again proved statistically nonsignificant.

*Reversal.*—A simple analysis of variance applied to the trials to criterion data showed the group means to differ significantly ($F = 5.45$; $df = 3/36$; $P < .01$). The group means are presented in the second row of Table 1, and it may be seen there that the results of this experiment are in close agreement with those of Exp. I. Group EI once again reversed significantly more slowly than any other group ($P < .005$ for all pairwise $t$ tests). Again Group CD had the lowest mean number of trials to reversal criterion, but not significantly so. And once again the overlearning effect was not obtained.

Although the detrimental effect on reversal learning of forced overlearning trials to the incorrect side has been verified, the overlearning effect remains elusive. Only one reasonable possibility occurred to us as a likely explanation, namely, that the number of overlearning trials, although at least equal to that employed in other reported studies, may have been insufficient. Thus, Exp. III was initiated in which the number of overtraining trials was increased from 160 to 300.

## EXPERIMENT III

### Method

*Subjects and apparatus.*—The Ss were 39 experimentally naive albino rats (18 males and 21 females) 90–95 days of age at the start of the study. Opaque doors were again used in the Y maze apparatus.

*Procedure.*—This experiment followed closely Exp. I, with the following exceptions:

(a) a 57-sec. intertrial interval was employed; (b) 300 overlearning trials were given (20 per day); (c) the acquisition and reversal criteria were made more stringent by requiring that the last 10 of the criterional trials be correct.

Since the acquisition-reversal interval was approximately the same as employed in the previous experiment, Group CD was eliminated from the present study. Thus, Groups C, EC, and EI contained 13 Ss each, assigned in the usual manner.

## Results and Discussion

*Acquisition and overlearning.*—The mean numbers of free trials to acquisition criterion were 61.7, 62.7, and 61.7 for Groups C, EC, and EI, in that order. On the 240 free overlearning trials Groups EC and EI averaged 2.9% and 4.1% errors, respectively. The difference between the group means was not significant.

*Reversal.*—The mean numbers of trials to reversal criterion are presented in the third row of Table 1. The results closely parallel those of Exp. I and II. A simple analysis of variance applied to the trials to criterion data produced an $F$ of 5.42 ($P = .01$). Group EI was significantly inferior to Group EC ($P < .02$) and to Group C ($P < .005$). Once again the overlearning effect failed to occur. (The Group C vs. Group EC comparison falls far short of significance.)

The results of the three experiments are consistent in demonstrating that severe retardation of reversal learning can be induced by a relatively small amount of forced incorrect responding. The marked inferiority of the EI groups in reversing suggests that avoidance of the incorrect place, or of the negative stimulus, is far from asymptotic at the time $S$ achieves criterion and reasonable habit mastery. Apart from theoretical considerations, one practical implication of this fact is that in a choice situation habit persistence can possibly be augmented much more efficiently by forced incorrect responding than by the more

natural method of permitting a moderate amount of free overtraining.

Our consistent failure to find the overlearning effect is difficult to understand in view of the relatively large number of reported studies in which the overlearning effect was obtained. It is not likely that the absence of the overlearning effect can be attributed to an excessive rate of incorrect responding in Group EC, as might be deduced from our interpretation of the overlearning effect. In Exp. III, Ss of Group EC averaged only 2.9% errors during overlearning; furthermore, analysis of the individual error scores showed that speed of reversal was not strongly related either to errors made during overlearning or to the sum of acquisition and overlearning errors.

A major difference between the present experiments and those in which an overlearning effect was reported lies in our use of forced trials during acquisition and overlearning. In Exp. IV the possible effect of this variable in the preceding studies was evaluated by eliminating forced trials during the acquisition and overlearning phases. In addition, the number of postcriterion trials was extended to a maximum of 800.

## EXPERIMENT IV

### Method

*Subjects and apparatus.*—The Ss were 36 experimentally naive female albino rats, 115–150 days of age at the start of the study. The Y mazes with opaque doors were again used.

*Procedure.*—Except for the elimination of forced trials and the number of postcriterion trials employed, the procedure closely paralleled that of Exp. III. The Ss were trained to criterion on a right-turning response and placed on reversal learning after 0, 200, 400, or 800 postcriterion trials (9 Ss per group). Twenty trials per day were given through all phases of the experiment, so that the 800 group was on overlearning for 40 consecutive days. All Ss were given a minimum of 7 days on reversal learning.

### Results

*Acquisition.*—The mean numbers of trials to acquisition criterion for the groups having 0 (the control

group), 200, 400, and 800 postcriterion trials were, in order, 48.9, 50.8, 51.6, and 54.6.

*Reversal.*—The mean numbers of trials to reversal criterion were, in the same order, 58.7, 111.7, 91.1, and 103.7. A simple analysis of variance showed the group means not to differ significantly ($F = 2.18$; $df = 3/32$; $.10 < P < .15$). A trend analysis based on the data of the first 7 reversal days indicated that the groups did not differ significantly with respect to the sum of correct responses made over the 7 reversal days. The differences between the groups' linear trends, as well as between their quadratic trends, were also nonsignificant.

## DISCUSSION

The results of Exp. IV are in accord with those of Exp. I, II, and III and restrict the range of conditions over which the overlearning effect can be expected. Although one may maintain that the 800 postcriterion trials were insufficient to obtain the overlearning effect, it should be noted that this number of overlearning trials is vastly greater than the numbers used in previous studies that have reported the effect. As to the question of the disparity between the present results and those of Pubols (1956), we can only offer the suggestion that the difference may be related to the fact that extramaze cues, apparently abundantly available in the latter study, are totally lacking in our apparatus. However, this factor could not easily explain the appearance of the overlearning effect in the "response learning" groups of the Brookshire, Warren, and Ball (1961) study.

It is unlikely that the absence of the overlearning effect in the present studies could have been predicted from the recent suggestion (e.g., Birch et al., 1960) that the effect may be attributable to a nonmonotonic relation holding between number of acquisition trials and resistance to extinction. According to this interpretation, reversal is faster after overlearning simply because overtraining leads to faster extinction of the approach response or, more generally, of the original habit. This hypothesis, it may be of interest to note, was first proposed by Jackson (1932) in what was most likely the first overlearning study.

Apart from the problem of handling the results of the present experiments, the hypothesis of nonmonotonicity between amount of acquisition training and resistance to extinction apparently does not generally describe the relationship between acquisition level and resistance to extinction, although there exists some confirming evidence (Murillo & Capaldi, 1961; North & Stimmel, 1960; Senko, Champ, & Capaldi, 1961). The major counterevidence comes from studies employing the bar pressing response in a simple instrumental setting (e.g., Harris & Nygaard, 1961; Margulies, 1961). In our own laboratory, using a continuous reinforcement schedule, we have found resistance to extinction of the bar pressing response to *increase* over groups having 278, 995, and 7070 mean numbers of acquisition responses, distributed over 2, 3, and 21 days, respectively.[2]

We are apparently faced with the following situation. Where the overlearning effect occurs, one can logically attribute the more rapid reversal of the overtrained *S*s to faster extinction of the formerly correct response, and there even exists some independent evidence that in certain situations this may be the correct interpretation (Birch et al., 1960). Where the overlearning effect fails to occur, as in the present experiments, it is of course conceivable that the failure is assignable to the *absence* of a nonmonotonic relation between amount of acquisition training and resistance to extinction, i.e., within the latter experimental conditions acquisition level and resistance to extinction enjoy the classical monotonic relationship. As already indicated the available evidence, even within the confines of an

[2] The research was conducted by F. Sperber and S. Gillman.

instrumental response, is not unambiguous. The North and Stimmel (1960) study is the only one known to the writers in which extinction of a simple instrumental response (starting time in a runway) was found to be nonmonotonically related to amount of acquisition training. In the Murillo and Capaldi (1961) and the Senko, Champ, and Capaldi (1961) studies S was required to guess whether or not a piece of cloth was present in a covered well by responding "in" or "out." Their "extinction" trials (cloth no longer present in well) were really reversal trials since S was reinforced for responding "out." Thus, in reality they were dealing with the effects of overtraining on reversal learning rather than on extinction, and it is only inferentially that their results can be claimed as support for the hypothesis of nonmonotonicity between amount of training and resistance to extinction.

It seems apparent that if the overlearning effect is to be explained in terms of an underlying relationship between amount of acquisition training and resistance to extinction, the variables influencing the latter relationship will have to be well specified before it can serve as a useful explanatory mechanism. Most probably these variables can best be isolated and studied within the context of a simple instrumental response.

## SUMMARY

Four experiments were conducted involving extensive overtraining of a position discrimination habit in rats. In Exp. I, II, and III, reversal learning of the position response was consistently and markedly retarded in those Ss that had a moderate proportion of their postcriterion trials forced to the incorrect side. In all three experiments, however, Ss that had the same proportion of postcriterion trials forced to the correct side did not show the overlearning effect, i.e., they did not reverse faster than control Ss that received no overlearning experience. In Exp. IV, run with all free trials, the overlearning effect again failed to appear, although the number of postcriterion trials was increased to a maximum of 800.

## REFERENCES

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, 60, 36–40.

BROOKSHIRE, K. H., WARREN, J. M., & BALL, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, 54, 98–102.

BRUNER, J. S., MANDLER, J. M., O'DOWD, D., & WALLACH, M. A. The role of overlearning and drive level in reversal learning. *J. comp. physiol. Psychol.*, 1958, 51, 607–613.

BUNCH, M. E. Transfer of training in the mastery of an antagonistic habit after varying intervals of time. *J. comp. Psychol.*, 1939, 28, 189–200.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, 59, 254–260.

D'AMATO, M. R., & JAGODA, H. Analysis of the role of overlearning in discrimination reversal. *J. exp. Psychol.*, 1961, 61, 45–50.

HARRIS, P., & NYGAARD, J. E. Resistance to extinction and number of reinforcements. *Psychol., Rep.*, 1961, 8, 233–234.

JACKSON, T. A. General factors in transfer of training in the white rat. *Genet. psychol. Monogr.*, 1932, 11, 1–59.

MARGULIES, S. Response duration in operant level, regular reinforcement, and extinction. *J. exp. Anal. Behav.*, 1961, 4, 317–322.

MURILLO, N. R., & CAPALDI, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, 61, 345–349.

MUROFUSHI, K. The effect of cue on reversal learning after varying periods of rest. *Psychologia*, 1957, 1, 37–46.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, 6, 227–234.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, 49, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, 46, 107–112.

SENKO, M. G., CHAMP, R. A., & CAPALDI, E. J. Supplementary report: Resistance to extinction of a verbal response as a function of the number of acquisition trials. *J. exp. Psychol.*, 1961, 61, 350.

STEVENSON, H. W., & WEIR, M. W. Response shift as a function of overtraining and delay. *J. comp. physiol. Psychol.*, 1959, 52, 327–329.

# THE VON RESTORFF ISOLATION EFFECT WITH MINIMAL RESPONSE LEARNING [1]

ARTHUR R. JENSEN

*University of California*

H. von Restorff (1933) found that when an "isolated" or perceptually emphasized item was included in a list of relatively homogeneous items, Ss would learn the isolated item quickly as compared with nonisolated items. The experimental literature and the numerous theoretical attempts to explain this phenomenon have been reviewed in a series of articles by Newman and Saltz (1958; Saltz & Newman, 1959, 1960). They concluded that the primary effect of isolation is to accelerate the learning of the isolated item as a response (Saltz & Newman, 1959, p. 450). Their experiments showed that the isolated item was indeed learned more rapidly as a response than a non-isolated item in the same serial position; the isolated item was emitted more frequently and occurred more often as an intrusion.

Learning a serial list of words or nonsense syllables, as in the Saltz and Newman studies, involves two phases: response learning and learning the serial order of the items. Thus, we may ask whether or not the isolation effect is manifest in the serial learning phase as well as in the response learning phase. The facilitation of response learning, emphasized by Saltz and Newman, is possibly only one result of isolation, and by itself may be inadequate to explain the total phenomenon.

The present experiment examined the isolation effect under conditions

in which (a) all the items in the list were already known to S so that all he had to learn was their serial order, and in which (b) S need not make a different response in the isolated than in the nonisolated condition, i.e., the isolated and nonisolated lists were identical in terms of the responses S was required to make.

## METHOD

*Subjects.*—Twenty men and 20 women were recruited from an introductory course in educational psychology.

*Procedure.*—To eliminate or at least minimize response learning, the serial lists in the experimental and control conditions were composed of nine colored geometric forms: triangles (T), circles (C), and squares (S) colored red (R), yellow (Y), and blue (B). Each shape appeared once in each of the three colors; stimuli of the same shape or color were never adjacent to each other in the list. The nine-item series was always preceded by three small white dots which served as the signal for anticipating the first item.

The stimuli were automatically projected onto a ground glass screen 2 ft. square. The figures were approximately 4 in. in size on the screen and the colors were vivid. The rate of presentation was 3 sec. per item, with a 6-sec. intertrial interval. The S sat approximately 10 ft. directly in front of the screen.

The Ss were tested individually. The experimental and control groups were given identical instructions. The S was told he would have to learn to a criterion of one perfect trial the order in which nine stimuli repeatedly appeared on the screen. The stimuli were named for S, who was then asked to repeat the names, e.g., RED TRIANGLE, etc. All Ss were easily able to give all the necessary responses before beginning the serial learning. The Ss learned by the anticipation method, responding by saying RED SQUARE, etc. They were urged to begin guessing on the very first trial and to guess when in doubt on subsequent trials.

## TABLE 1

### Summary Data for Control and Experimental Groups

| Group | Trials for Mastery of List | | Percentage Errors at Position 6 | | Order of Learning Position 6 | | Percentage Intrusions of Item 6 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| C | 22.75 | 7.58 | 15.92 | 3.25 | 7.50 | 1.62 | 22.69 | 9.65 |
| E | 23.30 | 9.87 | 10.31 | 3.63 | 4.55 | 2.35 | 18.96 | 7.12 |
| | $t < 1$ | | $t = 5.02^*$ | | $t = 4.50^*$ | | $t = 1.36$ | |

* $P < .001$.

*Experimental conditions.*—The order of the stimuli for Group C (Control) ($N = 20$) was: RS, YC, BS, YT, RC, BT, YS, RT, BC.

Previous experiments have shown that in a nine-item list Position 6 is generally the most difficult to learn. Therefore in the present experiment the sixth item was "isolated" or emphasized in Group E (Experimental) ($N = 20$). The rest of the list was the same as that learned by Group C. For Group E, instead of an actual blue triangle in Position 6, the words BLUE TRIANGLE appeared, printed in letters 2 in. high on the screen. Thus, Group E was required to learn the same responses as Group C; only the stimulus properties of the item in Position 6 differed for the two groups. The names of the shapes and colors, which are high frequency words in the Thorndike-Lorge word count, are probably so high in terms of response availability that it seems safe to assume there would be no appreciable difference in the strength of the naming response to the actual blue triangle and to the words BLUE TRIANGLE, especially none that would be evident under the 3 sec. anticipation time allowed in the present experiment.

### RESULTS AND DISCUSSION

The results are summarized in Table 1 and Fig. 1. The serial-position curves in Fig. 1 were obtained by determining each S's percentage of errors at each position and averaging these percentages for each group. Though response learning per se was practically eliminated by the method of the present experiment, the isolation effect was clearly manifested, the difference between Groups E and C in percentage of errors at Position 6

being significant ($P < .001$). The large percentage of errors at Position 7, immediately following the isolated item, contradicts the idea that isolation has the effect of breaking the list into two parts, each of which may be learned as a single list. This finding agrees with the conclusion of Newman and Saltz (1958) that the more rapid learning of the isolated item does not increase its effectiveness as a stimulus for eliciting the next item in the series.

As can be seen in Table 1, the groups did not differ significantly in the number of trials required to learn the list, which is also what Newman and Saltz (1958) found. Unlike the
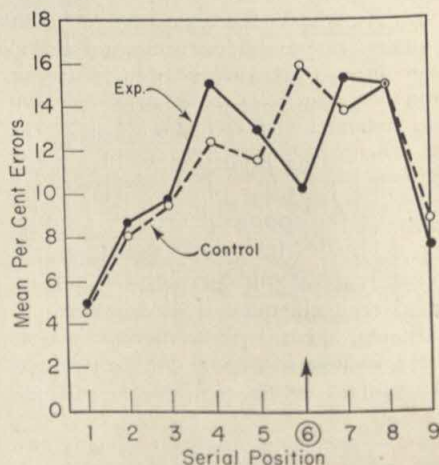


FIG. 1. Serial-position curves showing the isolation effect at Position 6 for Group E.

Newman and Saltz data, however, the conditions of the present study produced no significant differences between Groups E and C in the percentage of intrusions of Item 6 as an error in other positions. Saltz and Newman (1959) found that the isolated item was more likely to be emitted on the second trial, i.e., after a single presentation of the list. In the present study the total frequency with which BLUE TRIANGLE was given as a correct response on Trial 2 in Groups C and E was 2 each. The frequencies of BLUE TRIANGLE as an incorrect response on Trial 2 in Groups C and E were 10 and 13, respectively. The difference is non-significant.

The positions in the list were ranked for each $S$ in the order that $S$ learned them. The rank of a position was based on the number of the trial on which the last error occurred for that position. As shown in Table 1, the groups differed significantly in the mean rank order of learning Position 6. It has been found that when the items of a serial list are ranked in the order in which they are learned, the increment in errors on each item is a constant proportion of the total errors for all items (Jensen, 1962). A corollary is that for a given $S$ the same number of trials (or reinforcements) is required to learn each item, once the previous item in the order of learning has attained the criterion of learning. In other words, it appears that all the items in a serial list are of equal difficulty as regards the learning of their serial positions. Since all the items cannot be learned in one trial (unless the whole list is within $S$'s immediate memory span) they are necessarily learned in a particular order. The serial-position curve would result from the high degree of unanimity among $S$s in the order of learning the items. Though isolation changes the order of learning, so that the isolated item is learned sooner, it does not seem to be any easier in relation to the previously learned item

(regardless of its position) than is a nonisolated item. The differences between the percentage of errors on Item 6 and the previously learned item were 2.43 and 2.69 for Groups C and E, respectively. The hypothesis that isolation of an item changes only its order of being learned but not its difficulty is consistent with the general finding that isolation does not facilitate the learning of the list as a whole.

## SUMMARY

The von Restorff isolation effect was examined under conditions which minimized the role of response learning. Forty $S$s learned by the anticipation method the serial order of nine colored geometric forms, all of which $S$s could readily recall before having to learn their serial order. All $S$s learned the same *responses*; only the stimulus properties of the isolated item differed in the experimental condition.

The isolation effect was clearly manifested, showing fewer errors at the isolated position. The facilitation of response learning apparently is not the only effect of isolation and by itself cannot explain the total phenomenon. The number of intrusions of the isolated item did not differ significantly from that of the nonisolated item in the same position, nor did isolation facilitate learning the over-all list. It was suggested that when the effects of response learning per se are eliminated, isolation merely changes the order of learning the positions of the items in the serial list.

## REFERENCES

JENSEN, A. R. An empirical theory of the serial-position effect. *J. Psychol.*, 1962, **53**, 127–142.

NEWMAN, S. E., & SALTZ, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *J. exp. Psychol.*, 1958, **55**, 467–472.

SALTZ, E., & NEWMAN, S. E. The von Restorff isolation effect: Test of the intra-list association assumption. *J. exp. Psychol.*, 1959, **58**, 445–451.

SALTZ, E., & NEWMAN, S. E. Test of a "common sense" theory of the von Restorff effect. *Amer. Psychologist*, 1960, **15**, 451. (Abstract)

VON RESTORFF, H. Ueber die Wirkung von Bereichsbildungen im Spurenfeld. *Psychol. Forsch.*, 1933, **18**, 299–342.

(Received July 7, 1961)

# A COMPARISON OF REACTION TIME AND VERBAL REPORT
# IN THE DETECTION OF MASKED STIMULI [1]

ELIZABETH FEHRER AND IRVING BIEDERMAN

*Brooklyn College*

Fehrer and Raab (1962) found that reaction time (RT) to a visual stimulus was not affected when the stimulus was "masked" by subsequent stimulation of neighboring retinal areas (metacontrast). It was pointed out, however, that even when the test stimulus was phenomenally not present it did affect the appearance of the masking stimuli, since under these conditions the masks exhibited a readily detectable phi movement. In other words, the test stimulus, although masked as such, exerted both phenomenal effects, since its presence produced a change in the appearance of the masks, and behavioral effects, since it elicited as fast an RT as when it was presented alone.

The problem of the present experiments was to determine whether RT to a test stimulus would remain unaffected under masking conditions in which the presence of the test stimulus could not be detected phenomenally.

In the Fehrer and Raab study, the stimulus display consisted of three adjacent 2 × 2 in. light cells. The center square provided the test flash; the two flanking squares, the masking flashes. The luminance of all stimuli was approximately 18 ft-L.

A suitable testing condition for the present experiments required a test stimulus which was (a) sufficiently weak so that its presence could not be detected when the masking stimuli followed it, but also (b) sufficiently intense so that RT to it alone was faster than RT to the masking stimuli

plus the stimulus-onset delay ($\Delta t$). Unless this second provision obtains, it is not possible to determine whether a reaction to the combined stimuli has been initiated by the flashing of the test stimulus or by the masks. For example, if RT to the test stimulus when presented alone is 200 msec., and RT to the masks alone is 150 msec., then, at $\Delta t = 50$ msec., it is not possible to tell whether an RT of 200 msec. to the combined stimuli has been set off by the test or by the masking stimuli. At $\Delta t = 75$ msec., however, RTs averaging 200 msec. can be attributed to the test stimulus since here, if the reactions were to the masks, they would average 225 msec. (150 msec. plus 75 msec. for the $\Delta t$.)

These two requirements could not be fulfilled when both test and masking stimuli were brief light flashes. As the luminance of the test flash was reduced, RT to it became too long to meet the second requirement. In other words, at $\Delta t$'s at which phenomenal report failed, RT to the test stimulus alone was longer than RT to the masks plus the $\Delta t$.

A different condition, however, was found to be suitable. When the test stimulus consisted of a brief extinction of the otherwise constantly illuminated center square, then, when the masks (the two flanking squares) were flashed on, either simultaneously with the darkening of the center or slightly later, the darkening of the center square became impossible or very difficult to detect. Moreover, RT to the extinction of the test square alone was sufficiently rapid to

distinguish it from RT to the masks at certain $\Delta t$'s. It was this condition, therefore, that was explored in the present experiments.

## MAIN EXPERIMENT

### Method

*Subjects.*—The authors served as *Ss*. Both were well practiced in RT.

*Apparatus.*—The apparatus, described fully in a previous report (Fehrer & Raab, 1962), allows for the presentation of brief flashes of light or periods of darkness at specified stimulus-onset delays. The stimulus control panel and *E* were located in one room; the display and *S* in another.

The stimulus display consisted of three adjacent $2 \times 2$ in. light cells, each housing a cold cathode fluorescent lamp. The lamps were of the mercury vapor type, and coated with a calcium halo phosphor. When flashed, they yielded reasonably rectangular pulses of light. The center square area provided the test stimulus; the two flanking squares, the masking stimulus. The luminance of the stimuli, measured at a layer of diffusing tissue across the front of the light cells, was approximately 18 ft-L.

A 2-in. wide wooden frame, 18.5 in. long and 8.5 in. high, holding a sheet of polished glass, was pressed flat across the tissue covering the light cells. This frame was dimly illuminated at a level so low that it did not become visible until after 5 min. of dark adaptation.

The *S* was seated 15 ft. from the display. At this distance, the horizontal visual angle of the array was 2°.

Foveal fixation was employed throughout. Since the center square was constantly illuminated except for the brief test periods of darkness, no fixation point was required. The *S* was instructed to fixate the center of the test square when he heard the warning tone. Between trials, he looked at a dark area of the room.

The test stimulus throughout was a 5-msec. extinction of the light in the center square. The masking stimuli were 100-msec. flashes of light in the two flanking squares. The stimulus-onset delays studied were 0, 10, 20, 30, 50, and 75 msec.

RT trials were presented once every 15 sec. Each trial began with a warning tone which was followed by a 2.9-, 3.2-, or 3.5-sec. silent foreperiod. The three foreperiods were switch selected. RT was measured from the onset of the extinction of the light in the test square until *S* depressed a normally closed telegraph key. RTs to the test event alone and to the test-plus-masks could be read directly from the counter. RTs to the masks alone included the duration of the $\Delta t$, since, when no test stimulus was presented, the counter was started at the onset of the delay period.

*Procedure.*—Each *S* served in 30 experimental sessions, 5 at each of the six $\Delta t$'s. The first part of each session was devoted to RT; the second part to verbal report. Thirty-six RTs were measured at each session, 12 to the test stimulus alone, 12 to the masking stimuli alone, and 12 to the combination. The three stimuli were presented in random order and foreperiods were also randomized. The longest and the shortest RT of each 12 were discarded, yielding 10 RTs for each stimulus per session. Since five sessions were run at each $\Delta t$, each mean for each *S* is based on 50 RTs.

Each session began with 5 min. of dark adaptation after which the center light was turned on. Three practice RT trials preceded the recorded reactions. A 1-min. rest period followed the eighteenth reaction.

Verbal reports were obtained in the following way. First, *E* presented two examples of each of the two stimuli, correctly labeled. Thereafter, a total of 20 stimuli, either masks alone or test-plus-masks, were presented according to sequences determined by a table of random numbers. The interstimulus interval was 15 sec. The warning tone and the 2.9-sec. silent foreperiod preceded each presentation. After *S* judged which of the two stimuli had been presented, he was informed immediately whether he was correct. If *S* could not decide whether the test stimulus was present, he was allowed to request a repetition of the stimulus presentation. On the average, 1.8 such requests were made per session. Since five sessions were devoted to each $\Delta t$, there were 100 verbal reports at each delay. The sequence of sessions was counterbalanced in respect to $\Delta t$.

### Results

The mean RTs to the test, masking, and combined stimuli at each of the six $\Delta t$'s are shown in Table 1. The *SD*s (not shown) averaged 15.5, 10.4, and 11.7 msec., respectively. The percentages of correct verbal reports appear in the last column of the table. Better than chance ($P < .05$) detection is indicated by percentages of

TABLE 1

Mean Reaction Times and Percentages of Correct Verbal Reports Classified by Stimulus-Onset Delay ($\Delta t$)

| $\Delta t$ | S | Reaction Times | | | % Correct Reports |
|---|---|---|---|---|---|
| | | Test | Masks | Comb. | |
| 0 | IB | 170.0 | 159.4 | 162.9 | 63 |
| | EF | 182.3 | 156.7 | 155.9 | 53 |
| 10 | IB | 176.7 | 175.9 | 171.3[a] | 55 |
| | EF | 183.0 | 167.8 | 166.7 | 56 |
| 20 | IB | 180.7 | 185.0 | 174.9[a, b] | 68 |
| | EF | 179.5 | 176.0 | 167.9[a, b] | 55 |
| 30 | IB | 177.4 | 194.3 | 176.8[b] | 68 |
| | EF | 175.5 | 185.1 | 170.3[b] | 67 |
| 50 | IB | 183.9 | 215.0 | 180.1[b] | 83 |
| | EF | 182.7 | 207.2 | 179.8[b] | 88 |
| 75 | IB | 173.3 | 234.8 | 172.3[b] | 100 |
| | EF | 178.4 | 229.0 | 175.0[b] | 97 |

[a] Mean RT to combination significantly ($P < .05$) faster than RT to test stimulus.
[b] Mean RT to combination significantly ($P < .01$) faster than RT to masking stimuli.

60 or higher. (When $N = 100$, the standard error of 50% is 5%.)

It is readily apparent that the accuracy of the verbal reports improved with increase in $\Delta t$. This trend is, therefore, different from that found when both test and masking stimuli are equally luminous flashes of light presented in an otherwise dark room. In the latter situation, degree of masking (in this case, darkening) of the test flash increases with $\Delta t$ up to about 75 msec. The present experiment has shown that maximum masking of a brief period of darkness (as a result of the illumination of adjacent retinal areas) occurs with simultaneous presentation of test and masking stimuli. For 1 $S$, better than chance detection of the pulse of darkness required a $\Delta t$ of 30 msec. The other $S$ performed somewhat above chance level at all $\Delta t$'s other than 10 msec.

RT to the masks alone was faster than that to the test stimulus alone. The mask values as presented in Table 1 include the delay time. Net values for the purpose of this comparison, therefore, require subtraction of $\Delta t$.

At the $\Delta t$ of zero, the mean RT to the combined stimuli was virtually the same as RT to the masks and thus considerably faster than RT to the test stimulus. This indicates that the reaction to the combination was most probably initiated at the time of the presentation of the masks. The test stimulus had no clearly discernible effect.

Subject EF's reaction to the combined stimuli at $\Delta t = 10$ msec. also seems to have been initiated by the masks. Subject IB, on the other hand, reacted as fast to the test stimulus as he did to the masks plus the $\Delta t$. His RT to the combination, faster than that to either component, cannot be attributed to either stimulus alone.

At the $\Delta t$ of 20 msec., both $S$s reacted about as fast to the test stimulus as they did to the masks plus the $\Delta t$. RT to the combination, however, was reliably faster than that to either component alone. The facilitation at this $\Delta t$ and that for Subject IB at $\Delta t = 10$ msec. is similar to the facilitation in RT found at certain $\Delta t$'s in the study by Fehrer and Raab (1962). It is also similar to an intersensory (light and sound) facilitation reported recently by Hershenson (1962). The fact that this facilitation is sufficiently great to be significant (see Table 1) only at $\Delta t$'s at which both stimuli yield about the same RT is consistent with Hershenson's finding that maximum facilitation occurs when sound and light stimuli are separated by an interval equal to the difference in their respective RTs. Smaller amounts of facilitation are evident at all other delay intervals. In 11 out of the total 12 rows, RT to the combination was the fastest of the three.

The magnitude of the facilitation at $\Delta t = 20$ msec. is surprising in view of the fact that the test stimulus darkening was phenomenally a most inconspicuous event and required close attention for RT. RTs to the test stimulus, on the other hand, were reasonably fast. Speed of RT, rather than phenomenal appearance, may be the more important variable in facilitation.

The data at $\Delta t = 30$ msec. are the most relevant for the problem that initiated the present study, namely, the comparison of RT and verbal report in detection of the test stimulus. At this $\Delta t$, only 68% of the verbal reports were correct. RT to the combination, however, was not significantly different from RT to the test stimulus, but very significantly faster than RT to the masks plus the $\Delta t$. For each S, the distribution (not shown) of RTs to the combination was virtually the same as that to the test stimulus alone. There was no evidence of bimodality as there should have been if about one-third of the RTs to the combination had been set off by the masks. In addition, the SD of RTs to the combination was smaller than the SD for the test stimulus, 14.1 vs. 19.0 msec. for 1 S and 8.9 vs. 17.3 msec. for the other. The data, therefore, show that, even though careful observation failed in many cases to detect the presence of the test stimulus, RT apparently detected accurately at each stimulus presentation.

The same conclusion can be drawn from the results at $\Delta t = 50$ msec. It is apparent that RT to the combination was consistently initiated by the test stimulus. Verbal report was not entirely accurate.

## SUPPLEMENTARY EXPERIMENT

A brief experiment was run to determine whether the RT results of the main experiment might be due to the fact that both Ss knew of the presence of the test stimulus and, during all RT trials, were set to look for and react to the slight darkening. Since the sequence of stimuli was random, S never knew when this, a difficult event to detect, would occur. It seemed possible that naive Ss, knowing nothing of the test stimulus and therefore not being set for it, might not react more rapidly to the combination than to the masks alone.

Four naive undergraduate Ss reacted to the combination with a $\Delta t$ of 30 msec. and to the masks alone. The two stimuli were presented in random order with randomized foreperiods. Six sets of 36 RTs each were run with each S, two sets a day. Each set comprised 18 presentations of each of the two stimuli. The longest and the shortest RT of each 18 were discarded before the set means were computed. Since these Ss had had no practice in RT and exhibited very variable RTs on their first day, the first two sets of data were discarded. The final means, each based on 64 cases (four sets of 16 trials each), are presented in Table 2.

All Ss reacted reliably faster to the combination than to the masks, thus confirming the results of the main experiment. On the average, the naive Ss reacted 14.3 msec. faster to the combination than to the masks, a value very close to the mean difference of 16.2 msec. for the authors.

An attempt was made with 3 of the

### TABLE 2

MEAN REACTION TIMES AND PERCENTAGES OF CORRECT VERBAL REPORTS OF NAIVE Ss

| S | Reaction Times | | | % Correct Reports |
|---|---|---|---|---|
| | Masks | Combination | t | |
| 1 | 211.0 | 196.2 | 5.14 | — |
| 2 | 192.8 | 178.6 | 6.97 | 63 |
| 3 | 217.0 | 205.2 | 5.86 | 50 |
| 4 | 195.7 | 179.8 | 6.53 | 49 |

Note.—For combined stimuli, $\Delta t$ was 30 msec.

naive $Ss$ to determine whether they could learn to distinguish between the masks alone and the combination without informing them of the actual stimulus conditions. They were shown the two stimuli, labeled $a$ and $b$, five times each in alternation, and thereafter were asked to guess which condition was presented. After each report, they were informed whether they were correct. Four sets of 20 trials each were run, two sets a day on each of 2 days. The $S2$ discriminated somewhat better than chance ($P = .05$). The other 2 $Ss$ performed at chance levels. None reported any center darkening. The only differences they suggested were related to the duration of the masking stimuli.

## DISCUSSION

The data of these two experiments confirm and add to the results of two previous studies which showed that RT is determined by very brief changes in energy, and that therefore stimulus events originating later, though they are important in determining phenomenal characteristics, do not (except for certain facilitation effects) affect RT to the stimulus delivered first. It has already been pointed out that Fehrer and Raab (1962) showed that RT to a flash of light is not *increased* by subsequent light stimulation even though this results in the phenomenal darkening of the first flash. In another study (Raab, Fehrer, & Hershenson, 1961) it was shown that RT is independent of flash duration over the range of 10 to 500 msec., even though phenomenally the longer lasting stimuli are far brighter than the brief ones and might, therefore, be expected to elicit a faster RT. The present main experiment showed that RT can be initiated and determined by an event which is so successfully masked that it is often not detected by careful phenomenal observation. The supplementary experiment showed, further, that RT can be initiated

by an event whose presence is not even suspected by the reacting $S$.

It should, perhaps, be added that the present study falls in the general research area that includes the many recent experiments on subception, discrimination without awareness, etc. Eriksen's (1960) recent excellent review of many of these studies implies that behavioral indices, such as GSR, have proved to be, if anything, less sensitive in discrimination than well controlled phenomenal indices, such as forced choice. Our data, on the other hand, show that RT, a voluntary, objective response, can provide a more sensitive measure than verbal report in stimulus detection under masking conditions.

## SUMMARY

In the present study, we have compared the accuracy of two measures, reaction time and verbal report, in the detection of an event subjected to retroactive masking. A 5-msec. darkening of an otherwise steadily illuminated area was followed, after delays varying from 0 to 75 msec., by a 100-msec. illumination of two adjoining areas. At certain critical delays at which verbal detection of the test stimulus was little above chance accuracy, RT to the darkening of the test stimulus was not affected by the delayed presentation of the masking lights. Compared with verbal report, therefore, RT provided a far more accurate measure of the presence of the masked stimulus event.

## REFERENCES

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, **67**, 279–300.

FEHRER, E., & RAAB, D. Reaction time to stimuli masked by metacontrast. *J. exp. Psychol.*, 1962, **63**, 143–147.

HERSHENSON, M. Reaction time as a measure of intersensory facilitation. *J. exp. Psychol.*, 1962, **63**, 289–293.

RAAB, D., FEHRER, E., & HERSHENSON, M. Visual reaction time and the Broca-Sulzer phenomenon. *J. exp. Psychol.*, 1961, **61**, 193–199.

# DIFFERENTIAL EYELID CONDITIONING AS A FUNCTION OF THE CS-UCS INTERVAL [1]

THOMAS F. HARTMAN [2] AND DAVID A. GRANT

*University of Wisconsin*

The purpose of the investigation was to explore the relationship between conditioned discrimination and the CS-UCS interval. This relationship has not been systematically investigated, although Hilgard, Campbell, and Sears (1937) used intervals from 350 to 550 msec. and reported that intervals of 550 msec. produced better differential conditioning than did intervals of 400 msec. or less. This observation evidently led Hilgard and his co-workers to utilize 600- or 650-msec. CS-UCS intervals in subsequent differential conditioning (e.g., Hilgard, Campbell, & Sears, 1938; Hilgard, Jones, & Kaplan, 1951). Most other differential conditioning studies using the eyelid have involved a CS-UCS interval of approximately 500 msec., the interval optimal for simple eyelid conditioning (e.g., Spence & Beecroft, 1954; Spence & Farber, 1954).

A number of considerations might lead one to expect that the optimal CS-UCS interval for differential eyelid conditioning would be greater than the optimal CS-UCS interval for simple eyelid conditioning. Differential conditioning is more complicated than simple conditioning, and Hartman, Grant, and Ross (1960) point out that the limiting distribution for the latency of responses in the eyelid conditioning situation apparently is determined by some of the factors affecting reaction time, so that longer CS-UCS intervals might be appropriate in the differential conditioning situation. The work of Wickens et al. (e.g., Wickens, 1959) also implied that mediated responses to a stimulus complex may provide conditioned stimuli which will extend the usual optimal CS-UCS interval to a longer value. Certainly the use of the interval that is optimal for simple conditioning has led to relatively poor differential conditioning in earlier studies, and considerations such as those outlined above suggested that longer CS-UCS intervals might produce better conditioned discrimination.

## METHOD

*Apparatus.*—Except for the presence of two milk-glass windows for the presentation of the CS, the apparatus was essentially the same as that described by Hartman and Grant (1960). The two windows consisted of 10-cm. circular milk-glass disks with their centers horizontally 15 cm. apart. Ambient illumination was approximately 1 mL., and the CS consisted of an 0.8-mL. increase in brightness of one of the milk-glass windows. The UCS, a corneal air puff, was of 50-msec. duration, and its intensity was regulated by a 150-mm. column of mercury. The puff was sufficient to evoke a sharp reflex closure of the eyes of most of the Ss.

*Procedure.*—Each S was given 44 reinforced trials and 44 unreinforced trials. These were assigned randomly, balancing within blocks of 8 trials. For half of the Ss the right window was always reinforced, and for half the left window was always reinforced. The Ss were subdivided into four groups; one group receiving a CS-UCS interval of 400 msec., the others 600 msec., 800 msec., and 1000 msec. For all groups the CS duration

was 1100 msec. The intertrial intervals varied between 20 and 40 sec. with a mean of 30 sec. as programed by means of a Western Union tape transmitter.

Before each session Ss were given the "neutral" instructions used at the Wisconsin laboratory, and after 10 trials each S was interrupted and told that he should not aid or inhibit his natural eyelid responses.

*Subjects.*—The Ss were 56 women and 24 men who volunteered from classes in elementary psychology at the University of Wisconsin. Assignment of Ss to conditions was random within each replication of the eight experimental conditions until it became necessary to modify the assignment to equalize sexes in each group. Three Ss were discarded; 2 adapted to the UCS and 1 gave a record impossible to score because of a high random blink rate.

## RESULTS

Owing presumably to the low intensity stimuli and short duration UCS used in the present experiment the random blink rate was low enough so that CS-UCS interval effects were apparent without a correction for the scoring interval. Therefore all eyelid closures with latencies from 210 msec. to 10 msec. after the UCS onset were scored as anticipatory eyelid responses, giving scoring intervals of 200, 400, 600, and 800 msec., respectively, for the CS-UCS intervals of 400, 600, 800, and 1000 msec.

Figure 1 shows the mean percentage of anticipatory responses to the positive and negative stimuli plotted as a function of successive blocks of trials. The first block was four trials, subsequent blocks were eight trials each. Because the right or left position of the positive CS produced no difference in results, this factor is ignored in this and all other figures and computations. It is readily apparent from Fig. 1 that conditioned discrimination increased as the CS-UCS interval was extended from 400 through 600 to 800 msec. There is some diminution in conditioned
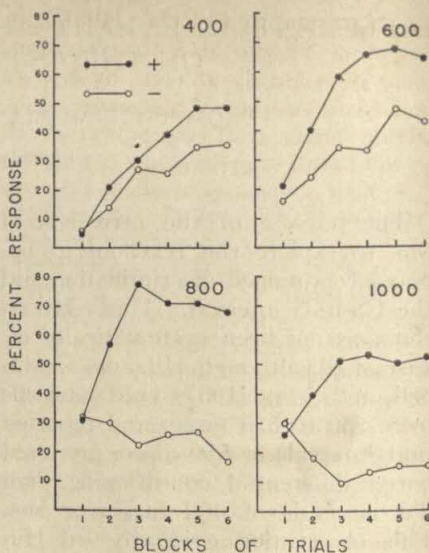


Fig. 1. Percentage frequency of anticipatory responses to the positive and negative stimuli during successive blocks of acquisition trials for the four CS-UCS intervals.

discrimination with the 1000-msec. CS-UCS interval. The standard error of the differences between pairs of positive and negative points in Fig. 1 ranged between 2.00 and 3.50 percentage units, so that for all groups there was statistically significant discrimination between the positive and negative stimuli in the last three blocks of trials. It is of interest to note that with the short CS-UCS intervals the responses to both the positive and negative stimuli increased during the training session, whereas with the longer CS-UCS intervals the responses to the negative stimuli actually decreased during the course of training. The greater conditioned discrimination obtained with the longer CS-UCS intervals is thus due to a lowered response to the negative stimulus.

On the last 32 trials Ss within each group were separated into two groups according to the criterion suggested

by Hartman and Ross (1961) for detecting Ss who give the voluntary form response discovered by Spence and his co-workers (Spence & Ross, 1959; Spence & Taylor, 1951). All Ss with time derivatives (dx/dt) of anticipatory responses to light greater than 35% of the time derivative of the reflex to the UCS on more than half of their responses were classified as voluntary responders (Vs). Other Ss were called conditioners (Cs). The derivative criterion classed 29 Ss and Vs; 7 in each of the 400-, 600-, and 1000-CS-UCS interval groups, and 8 in the 800-msec. group. There were thus 12 or 13 Cs in each group. The mean percentage of anticipatory closures to the positive and negative stimuli over the last 32 trials is plotted for both the Vs and Cs as a function of CS-UCS interval in Fig. 2. The difference between the positive and negative curves can be used as an index of the degree of discrimination. Noting these differences, it will be
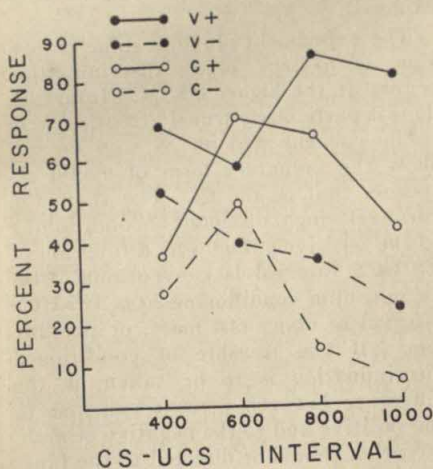
seen that degree of discrimination increases progressively as CS-UCS interval increases for the Vs. For the Cs, however, there appears to be an optimum degree of differential conditioning at the 800-msec. CS-UCS interval. Because the frequency of responses to the negative stimulus in the Cs approaches zero at the 1000-msec. interval and presumably would stay low at higher intervals and because the responses to the positive stimulus gradually decrease as the CS-UCS interval is extended beyond 800 msec., the difference between response rates to the positive and the negative stimuli may be expected to decrease with longer CS-UCS intervals.

Even without a correction for the scoring interval the Cs show the same response function to the positive stimulus that is usually obtained in simple conditioning (e.g., Kimble, 1947). Responses to the negative stimulus are also clearly affected by the CS-UCS interval in spite of the fact that these stimuli are never reinforced so that there is no real CS-UCS interval for the negative stimuli. This phenomenon involves some sort of transfer from the positive stimulus where CS-UCS interval has direct meaning.

The differences between the CS-UCS interval functions for the Vs and the Cs indicate that in the differential conditioning situation as in the simple conditioning situation (Gormezano & Moore, 1962; Hartman & Grant, 1962; Spence & Ross, 1959), the Vs follow different behavioral laws in the eyelid conditioning situation than do the Cs.

In view of the differences in the response functions between the Vs and Cs, one might legitimately ask why their data were combined in Fig. 1. The answer is that if the



FIG. 2. The percentage of anticipatory eyelid closures to the positive and negative stimuli for voluntary responders (V) and nonvoluntary responders (C) on the last 32 acquisition trials as a function of the CS-UCS interval.

TABLE 1

ANALYSES OF VARIANCE OF DIFFERENCES IN
FREQUENCY OF RESPONSES TO THE
POSITIVE AND NEGATIVE STIM-
ULI ON LAST 32 ACQUISI-
TION TRIALS

| Source of Variation | Voluntary Ss | | Conditioned Ss | |
|---|---|---|---|---|
| | df | F | df | F |
| Between CS-UCS intervals | (3) | 2.12* | (3) | 5.36** |
| Linear | 1 | 8.30** | 1 | 9.80** |
| Quadratic | 1 | 0.23 | 1 | 3.00 |
| Cubic | 1 | 1.03 | 1 | 3.27 |
| Error (MS) | 25 | (25.43) | 47 | (20.19) |

Note.—The between CS-UCS sum of squares was
computed from weighted means and is unbiased. In
the orthogonal decomposition of the trend unweighted
means were used to simplify the problem of unequal
numbers of cases, but the bias was negligible because
the variation in the numbers of cases per CS-UCS
interval group was never greater than 1.
* $P = .05$.
** $P = .01$.

curves of Fig. 1 were given separately
for Vs and Cs they would show es-
sentially the same acquisition func-
tions except for higher response level
in the Vs, especially to the positive
stimulus at the 1000 CS-UCS interval,
as shown in Fig. 2.

Table 1 summarizes the analysis
of variance of the difference scores
of responses to the positive and nega-
tive stimuli on the last 32 trials as
shown in Fig. 2. The significant
linear components of the trends of the
difference scores show that condi-
tioned discrimination increases with
increases in CS-UCS interval. Al-
though there was no significant quad-
ratic trend for the Cs, it seems evident
that maximum discrimination occurs
with CS-UCS intervals of 800 to
1000 msec.

Examination of the response la-
tencies for high and low discriminators
among both the Vs and the Cs at each
CS-UCS interval showed no consistent
latency differences between the high
and low discriminators. The Vs,
however, had consistently shorter

response latencies than the Cs in each
group. The latency differences be-
tween these two types of Ss increased
with increases in the CS-UCS interval.
When responses were made to the
negative CS, these responses did not
differ in form or latency from re-
sponses made to the positive CS.

DISCUSSION

Differential conditioning or condi-
tioned discrimination was found to
increase with increasing CS-UCS inter-
vals. This was due largely to the greater
reduction in responses to the negative
stimulus with the longer CS-UCS inter-
vals. Although there was no statis-
tically significant evidence for an opti-
mum degree of conditioned discrimina-
tion within the CS-UCS range studied,
there is some indication that there is such
an optimum for Ss who do not show the
voluntary form of eyelid response. There
may be a corresponding optimum CS-
UCS interval for conditioned discrimina-
tion among Ss showing the voluntary
form response, but it is probably at a
longer CS-UCS interval than those
explored.

The experiment provides some indica-
tion as to why better discrimination
occurs at the higher CS-UCS intervals.
This is particularly true if attention were
to be concentrated on Ss who do not
show the voluntary form of eyelid re-
sponse, that is, the Cs. For them the
curve relating the amount of conditioning
to the positive stimulus as a function of
CS-UCS interval is conventional with
an optimum conditioning at a CS-UCS
interval of about 600 msec. or possibly
less. If the measure of conditioned
discrimination is to be taken as the
difference between percent response to
the positive and to the negative stimuli,
attention should be directed to the func-
tion relating responses to the negative
stimuli to the CS-UCS interval, for the
greater discrimination at longer CS-UCS
intervals is due to the rapid drop in this
function. Actually it is not immediately
obvious why there should be a functional

relationship of this form for responses to the negative stimulus. There is no true CS-UCS interval for the negative stimulus as the UCS never follows the negative CS. Therefore the interval is defined by events on trials where the positive CS is given. To get a function to the negative stimulus requires that events associated with the positive stimulus affect responses to the negative function, but it requires some manipulation of this concept to say the least. Alternatively, some additional mechanism that cannot act at the shorter CS-UCS intervals may be involved in the longer CS-UCS intervals. This conjecture is made attractive by the fact that the acquisition curves to the negative stimulus in Fig. 1 progress downward at the 800- and 1000-msec. intervals and upward at the 400- and 600-msec. intervals.

The difference might depend upon perceptual or upon reaction systems or both. On the one hand, longer CS-UCS intervals might permit a more complete perceptual response to the positiveness or negativeness of the CS, and this perceptual response might provide the effective CS conditioning along the lines proposed by Wickens et al. (Wickens, 1959). On the other hand, inhibition of a response may simply require more time and be favored by longer CS-UCS intervals. Of course, a combination of both principles may be involved. Certainly longer CS-UCS intervals produce longer latency responses (Boneau, 1958). Also the Vs show generally shorter latency responses than do the Cs who inhibit more effectively. But there are no differences in latencies between the good and poor discriminators among either the Vs or the Cs. Actually the experiment provides little evidence for details of the mechanism of the better discrimination at the longer CS-UCS intervals, but the fact that the Cs give the conventional maximum for positive conditioning at about 600 msec. CS-UCS interval for responses to the positive stimulus may favor an interpretation in terms of inhibition requiring longer time rather than an interpretation based on a mediating perceptual response.

It should be noted that the Vs generally give more responses and are poorer at inhibition than are the Cs. In this respect their performance is like that reported by Hartman and Grant (1962) and also by Gormezano and Moore (1962), who note that the Vs show poorer extinction generally. As was pointed out by Hartman and Grant, the Vs are by no means uniformly appropriate in their responses to the stimuli. In some respects their performance is reminiscent of Pavlov's excitable type (Pavlov, 1927, pp. 285–300; 1928, pp. 360–390).

The results of the present experiment may provide a possible clue to the discrepancy between the results of Hilgard, Jones, and Kaplan (1951) who found less conditioned discrimination in high anxious Ss and the results of Spence and Farber (1954) and Spence and Beecroft (1954) who found greater discrimination in high anxious Ss. Hilgard, Jones, and Kaplan used a CS-UCS interval of 650 msec., whereas Spence and Farber and Spence and Beecroft used 500 and 490 msec., respectively. The latter may have dealt with simpler basic response principles, whereas the former may have had additional complicating inhibitory principles operating to reverse the anxiety discrimination relationship. Whether the difference in CS-UCS interval will turn out to be the basis of the anxiety discrepancy or not, CS-UCS interval certainly seems to account for the hitherto puzzling discrepancy in the published acquisition curves of responses to the negative stimulus. Hilgard, Jones, and Kaplan found responses to the negative stimulus to decrease with successive acquisition trials as did our Ss at CS-UCS intervals greater than 600 msec. Spence and Beecroft found increasing responsiveness to the negative stimulus as acquisition progressed. In this respect their findings were more like our 400- or 600-msec. CS-UCS interval groups. The responses to the negative stimuli in the Stanford and Iowa experiments were thus appropriate to the CS-UCS intervals utilized.

## SUMMARY

Differential eyelid conditioning was studied at four CS-UCS intervals (400, 600, 800, and 1000 msec.) with 20 Ss in each interval. All Ss received 88 training trials, 44 reinforced trials with the positive CS, and 44 unreinforced trials with the negative CS. The CS was a light and the UCS was a corneal air puff; the positive and negative CS appeared in two glass windows. The principal findings were as follows:

1. Conditioned discrimination increased as the CS-UCS interval was increased. For Ss who never or rarely showed the voluntary form of eyelid response (Cs) the amount of conditioning to the positive stimulus showed the conventional optimum at about the 600-msec. CS-UCS interval or less. The increased discrimination was caused by rapid decrease in the percentage of responses to the negative stimulus as the CS-UCS interval was extended.

2. There were indications of optimum discrimination in the Cs at about the 800-msec. CS-UCS interval, but the difference between responsiveness to the positive and negative stimulus increased progressively over the CS-UCS range studied for Ss who showed the voluntary response form (Vs).

3. Although it was conjectured that the superior differential conditioning at the longer CS-UCS intervals might have been due to a more complete mediating perceptual response, an interpretation in terms of longer time intervals required for inhibition also seemed plausible.

## REFERENCES

BONEAU, C. A. The interstimulus interval and the latency of conditioned eyelid responses. *J. exp. Psychol.*, 1958, 56, 464–471.

GORMEZANO, I., & MOORE, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, 63, 487–494.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, 60, 89–96.

HARTMAN, T. F., & GRANT, D. A. Effects of pattern of reinforcement and verbal information on acquisition, extinction, and spontaneous recovery of the eyelid CR. *J. exp. Psychol.*, 1962, 63, 217–226.

HARTMAN, T. F., GRANT, D. A., & ROSS, L. F. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rep.*, 1960, 7, 305–311.

HARTMAN, T. F. & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, 61, 334–338.

HILGARD, E. R., CAMPBELL, A. A., & SEARS, W. N. Conditioned discrimination: The development of discrimination with and without verbal report. *Amer. J. Psychol.*, 1937, 49, 564–580.

HILGARD, E. R., CAMPBELL, R. K., & SEARS, W. N. Conditioned discrimination: The effect of knowledge of stimulus-relationships. *Amer. J. Psychol.*, 1938, 51, 498–506.

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, 42, 94–99.

KIMBLE, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, 37, 1–15.

PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

PAVLOV, I. P. *Lectures on conditioned reflexes.* (Trans. by W. H. Gantt) New York: International, 1928.

SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, 48, 399–403.

SPENCE, K. W., & FARBER, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, 47, 127–134.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, 58, 376–385.

SPENCE, K. W., & TAYLOR, J. Anxiety and the strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, 42, 183–188.

WICKENS, D. D. Conditioning to complex stimuli. *Amer. Psychologist*, 1959, 14, 180–188.

# MIXING OF TWO TYPES OF S-R ASSOCIATIONS IN A CHOICE REACTION TIME TASK [1]

ROBERT E. MORIN        AND        BERT FORRIN

*University of Arizona*                    *University of Washington*

Recent studies by Brainard, Irby, Fitts, and Alluisi (1962) and Mowbray (1960) have shown that the time required to name a numeral is independent of the number of numerals in the stimulus set. Cast in the context of information theory, these results broaden the spectrum of tasks for which independence of transmitted information and reaction time (RT) has been found. Other investigators, all of whom used key pressing responses, have obtained similar results with extended practice (Mowbray & Rhoades, 1959), high stimulus-response compatibility (Leonard, 1959), and stimulus uncertainties of greater than 3 bits (Seibel, 1959).

The above findings are of interest, first, because they contradict what has been regarded as well-established empirical generalization. Since the work of Merkel (1885), it has been rather generally accepted that disjunctive RT increases as the number of stimulus alternatives is increased. In the past decade, with the application of information measures to RT tasks, the generalization has been refined to state that choice RT is a positive linear function of transmitted information.

The new results are of further interest because they imply a mode of information processing different from that suggested by earlier studies. If

choice RT is a positive linear function of transmitted information expressed in bits, it is easy to conceptualize the process of transmitting information as a *series* of binary decisions, each decision taking constant time. On the other hand, if there is independence of choice RT and transmitted information, the picture is one of a *parallel* processing mechanism with each S-R pair having its own "private line."

In view of the contrasting results of very recent and earlier studies, two types of S-R associations may be distinguished in terms of behavioral criteria. One type, hereafter called Type N (N for independence), includes associations which normally produce independence of choice RT and transmitted information (e.g., naming numerals). The second type, hereafter called Type D (D for dependence), includes associations for which RT has been found to be proportional to transmitted information (e.g., numeral responses to geometric symbols). All of the studies upon which these preliminary distinctions are based have used homogeneous sets of associations, either all Type N or all Type D pairs. The present investigation is concerned with the effects of mixing the two types of pairs in a single task. Each type of pair will serve as a context for the other. The fundamental question is whether or not the two types of pairs interact. Are RTs for Type N pairs affected by the inclusion of Type D pairs, and what effect, if any, does the mixing have on RTs for the latter pairs?

The answers to these questions should help to define the generality of recent findings showing independence of transmitted information and RT. Furthermore, they should extend the basis for theorizing concerning the translation mechanisms involved in information processing.

## METHOD

*Subjects.*—The *S*s were 50 male students recruited from classes in elementary psychology at the University of Texas. Participation in the experiment partially fulfilled a course requirement. Ten *S*s were assigned at random to each of five conditions. The *S*s were tested individually.

*Apparatus.*—The *S* was seated approximately 2 ft. in front of a milk glass screen behind which was located an automatic slide projector. The dimensions of the illuminated field were 4.0 × 5.5 in., and the height of the projected images, which varied in shape, ranged from 2.0 to 2.5 in.

The RTs were determined to the nearest .01 sec. by a Hewlett-Packard electronic counter, and were printed on paper tape by a Hewlett-Packard digital recorder. A verbal response by *S* activated a voice key which stopped the counter, terminated the projected image, and initiated an automatic slide changing operation. The interval between a response and the appearance of the subsequent slide was 2.1 sec. Errors were recorded by *E*.

*Procedure.*—After a brief period in which *S*s learned the S-R associations from an instruction card, 12 blocks of 32 trials were given. There was a 4-min. rest after Block 6 and shorter rests of approximately 30 sec. while *E* changed slide trays after other blocks. Stimulus sequences were constructed to equate both frequencies of occurrence of all stimuli within a condition and first-order transitional probabilities. The *S*s were instructed to react as rapidly as possible and to keep errors at a minimum.

*Experimental conditions.*—The five experimental conditions are described in Fig. 1. For purposes of exposition, the conditions have been divided into two groups, Exp. A (Cond. I, III, and IV) and Exp. B (Cond. II, III, and V).

In Exp. A, RTs to the critical Type N pairs (pairs common to all conditions of the experiment) were studied as a function of both the presence and type of associations

| EXPERIMENT A | | | EXPERIMENT B | | |
|---|---|---|---|---|---|
| COND. | CRITICAL PAIRS | CONTEXT | COND. | CRITICAL PAIRS | CONTEXT |
| I | 2 - 2<br>8 - 8 | | II | + - 4<br>■ - 7 | |
| III | 2 - 2<br>8 - 8 | + - 4<br>■ - 7 | III | + - 4<br>◪ - 7 | 2 - 2<br>8 - 8 |
| IV | 2 - 2<br>8 - 8 | 4 - 4<br>7 - 7 | V | + - 4<br>■ - 7 | ● - 2<br>▲ - 8 |

FIG. 1. Schematic diagram of the experimental conditions.

used as context. In all conditions of this experiment, the critical stimuli were 2's and 8's; the correct response was to name the numeral displayed. In Cond. I only the critical pairs were employed; they were without context. In Cond. III two Type D pairs provided the context for the critical Type N pairs. The stimulus set was increased by the inclusion of crosses and squares to which the correct responses were 4 and 7, respectively. In Cond. IV two Type N pairs (4-4, 7-7) were used as context for the critical pairs.

In Exp. B, RTs to the critical Type D pairs (cross-4, square-7) were investigated under conditions of no context (Cond. II), a context of Type N pairs (Cond. III), and a context of Type D pairs (Cond. V). It is to be noted that data of Cond. III are included in both experiments.

*Response measures.*—Four response measures were computed for each *S*: (*a*) mean RT to the critical stimuli; (*b*) information transmitted by the total set of stimulus elements $(T_t)$; (*c*) information transmitted by the critical stimuli $(T_c)$; and (*d*) rate of information transmission for the critical stimuli $(T_c/RT)$.

In Cond. III, IV, and V all response measures but the second were based on the 192 trials on which critical stimuli (2's and 8's in Exp. A, crosses and squares in Exp. B) were presented. In Cond. I and II, since critical stimuli were presented on all trials, two alternative procedures were possible: (*a*) to determine response measures from the data of all 384 trials, and (*b*) to use only the initial 192 trials. The first alternative serves to equate total amount of practice on the task for all conditions; the second provides for comparable amounts of practice in responding to the critical stimuli. In the case of the three response measures in question, both procedures were followed with nearly

identical results. Decisions concerning differences among experimental conditions were not affected by the number of trials upon which response measures for Cond. I and II were based. For this reason all statistical analyses reported involve the comparison of conditions with total practice on the task equated.

As noted above two measures of transmitted information were calculated for the experimental conditions. The first, $T_t$, was based on responses to all stimuli (384 trials) and is the measure commonly employed in studies relating choice RT to transmitted information. The second, $T_c$, was based solely on responses to critical stimuli. The latter measure was required because $T_t$ does not discriminate between errors to critical and to contextual stimuli. $T_t$ and $T_c$ were identical in Cond. I and II. The maximum values for $T_t$, attainable with errorless performance, were 1 bit in Cond. I and II and 2 bits in Cond. III, IV, and V. The maximum value for $T_c$ was 1 bit for all conditions.

## RESULTS

Means for each of the response measures are given in Table 1. For three measures—RT, $T_c$, and transmission rate—overall comparisons of the means for the three conditions within each experiment were made. When an analysis yielded a statistically significant $F$, Duncan's range test was applied to make more analytical comparisons. Levels of significance cited below are for Duncan's test.

In Fig. 2 mean RTs for the conditions of both experiments are plotted



FIG. 2. Reaction time in seconds plotted as a function of total transmitted information in bits.

as a function of $T_t$. As is apparent from the figure, the number of trials on which response measures for Cond. I and II are based does not materially alter the nature of the RT-$T_t$ functions.

Initially the results may be examined to determine whether the associations used met the behavioral criteria for Type D and Type N pairs. The significant slope of the line defined by the means for Cond. II and V ($P < .001$) satisfies the requirement for the identification of symbol-numeral associations as Type D pairs. For the numeral-numeral associations (Cond. I and IV) RT was independent of mean $T_t$. In fact, the observed slope, though nonsignificant ($P > .05$), was slightly negative. However, fewer errors to the critical stimuli in Cond. IV than in Cond. I resulted in a significantly higher $T_c$ ($P < .01$) and a significantly faster transmission rate ($P < .05$) in the former condition. Though these findings indicate that the two conditions were not behaviorally equivalent in all respects, it is nevertheless clear that perform-

TABLE 1

MEAN RESPONSE MEASURES FOR THE
CONDITIONS OF EXP. A AND B

| Exp. | Cond. | RT (Sec.) | $T_t$ (Bits) | $T_c$ (Bits) | $T_c$/RT (Bits/ Sec) |
|---|---|---|---|---|---|
| A (Type N pairs) | I | .52 | .93 | .93 | 1.81 |
| | III | .60 | 1.92 | 1.00 | 1.68 |
| | IV | .49 | 1.97 | .98 | 2.04 |
| B (Type D pairs) | II | .59 | .86 | .86 | 1.48 |
| | III | .71 | 1.92 | .86 | 1.22 |
| | V | .72 | 1.84 | .93 | 1.32 |

ance in Cond. IV was not inferior to that in Cond. I.

The effects of mixing the two types of pairs can be determined by comparing performance in Cond. III with that in other conditions of each experiment. In Exp. A the mean RT for Cond. III was significantly greater than the means for Cond. I ($P < .01$) and IV ($P < .001$). Reactions to numeral stimuli (2's and 8's) were slowed by the addition of geometric symbols to the stimulus set. The increase in RT in Cond. III was accompanied, however, by a somewhat compensating decrease in error rate to the critical stimuli such that mean transmission rate, though lower than that for both other conditions, differed significantly only from Cond. IV ($P < .001$).

In Exp. B reactions to the critical stimuli (crosses and squares) were similarly affected whether numerals or geometric symbols were used as contextual stimuli. Conditions III and V did not differ significantly on any response measure. As compared to Cond. II, however, performance in Cond. III was characterized by a significantly higher mean RT ($P < .001$) and a significantly lower mean transmission rate ($P < .001$).

## DISCUSSION

The results lend support to the generalization that RT is independent of mean transmitted information in a numeral-naming task. The response to a particular numeral is not degraded by the fact that other numerals might have occurred. Thus, numeral-numeral associations appear to be functionally isolated from one another.

A test of the generality of this conclusion is provided by the mixing of Type N and Type D pairs. If Type N associations are in functional isolation not only from one another but from Type D associations as well, RTs for

Type N pairs should be unaffected by the inclusion of Type D pairs in the S-R set. Complete functional isolation of Type N pairs would further imply that the presence of such pairs as context should exert no influence on reactions to Type D pairs. If the translation process involved in the naming of a numeral is an encapsulated and automatic event, Ss should, so to speak, be able to let the numeral-numeral associations take care of themselves; reactions to crosses and squares in Cond. III should be much as they were in Cond. II.

Neither of the outcomes consistent with the extension of the concept of functional isolation to the case of mixed pairs was obtained. Response latencies for Type N pairs were lengthened by the addition of Type D pairs. Though increased RTs were partially compensated for by a decrease in errors it is apparent that numeral naming was sensitive to contextual associations. In addition, RTs for Type D pairs were clearly related to the bivariate probability distribution of *all* stimulus-response events. A context of Type N pairs (Cond. III) produced increases in RT of the same magnitude as a context of Type D pairs (Cond. V). The Ss in Cond. III did not react as though crosses and squares were the only stimuli which might occur.

The present results indicate the RTs for numeral-numeral pairs can be affected by the context in which they are imbedded. The disruption of functional isolation for Type N pairs suggests that an elementary parallel processing model is too simple for characterizing translation mechanisms, even for highly overlearned associations.

## SUMMARY

Two types of S-R associations were distinguished in terms of behavioral criteria. Type N associations (e.g., numeral-numeral pairs) normally produce independence of choice RT and transmitted information. For Type D associations (e.g., symbol-numeral pairs) choice RT is a positive linear function of transmitted information. The present study investigated the effects of mixing

Type N and Type D associations upon the RT-transmitted information function.

Ten male Ss were randomly assigned to each of five experimental conditions defined by the character of the S-R set: (I) two Type N pairs, (II) two Type D pairs, (III) two Type N and two Type D pairs, (IV) four Type N pairs, and (V) four Type D pairs. Analyses examined the reactions to the two critical Type N pairs common to Cond. I, III, and IV (Exp. A) and to the two critical Type D pairs common to Cond. II, III, and V (Exp. B). In Exp. A, mean RT to the critical Type N pairs was not degraded by the expansion of the S-R set from two to four Type N pairs (Cond. I vs. Cond. IV); in contrast, mean RT was increased significantly by the addition of two Type D pairs (Cond. I vs. Cond. III). In Exp. B, response latencies to the critical Type D pairs were shortest in Cond. II. The addition of two Type N pairs to the S-R set (Cond. III) lengthened mean RT to the critical Type D pairs by an amount comparable to that obtained by the addition of two Type D pairs (Cond. V).

The results of this study were considered to limit the generality of the proposition that RT for numeral-numeral associations is independent of the informational properties of the task.

## REFERENCES

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

LEONARD, J. A. Tactual choice reactions: I. *Quart. J. exp. Psychol.*, 1959, 11, 76–83.

MERKEL, J. Die zeitlichen Verhältnisse der Willensthatigkeit. *Phil. Stud.*, 1885, **2**, 73–127.

MOWBRAY, G. H. Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, 1960, **12**, 193–202.

MOWBRAY, G. H., & RHOADES, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

SEIBEL, R. Discrimination reaction time as a function of the number of alternatives and of the particular stimulus-response patterns. *Amer. Psychologist*, 1959, **14**, 396. (Abstract)

# HUE GENERALIZATION AND HUE DISCRIMINABILITY
# IN *MACACA MULATTA* [1]

## LEO GANZ [2]

*University of Chicago*

The principle that when an organism has difficulty in discriminating two stimuli, he will be more likely to generalize a response from the one, a training stimulus (CS), to the other, a generalization stimulus (GS), is widely adopted by behavior theorists (Brown, Bilodeau, & Baron, 1951; Gewirtz, Jones, & Waerneryd, 1956; Hull, 1943; Lashley & Wade, 1946; Pavlov, 1927).

In the context of such unanimity of opinion, the paucity of corroborative material is surprising. The experimental analysis of this question thus far has concentrated on generalization gradients to hue in different spectral regions where it is known that sensitivity varies. In the pigeon these gradients have failed to reflect what we know of that $S$'s differential hue threshold function (Guttman & Kalish, 1956). A similar study in which intra-$S$ gradient comparisons were possible did obtain consistent differences in generalization related to wave length, but not the expected relationship to the jnd function (Blough, 1961). In the human, generalization gradients obtained from voluntary responses (method of single stimulus) have parallelled the jnd

function (Kalish, 1958). The question remains, clearly, very much open.

The evidence that is available to us suggests that the color vision of the rhesus macaque (Grether, 1939) bears a strong resemblance to human trichromacy. This makes it feasible to compare, as is done in the present study, the generalization gradients of individual rhesus macaques in a succession of different spectral regions with the corresponding human differential hue threshold function. Our working hypothesis is that in spectral regions where the jnd, in millimicrons of wave length, is small (good discriminability) the generalization gradient will be steep in slope for CS and GS contained in that region; where the jnd is large (poor discriminability) the gradient will be shallow.

## METHOD

*Subjects.*—The $Ss$ were 6 pre-adolescent *Macaca mulatta*. Two $Ss$, Ka and Bu had been used in previous studies involving a black-white discrimination for solid food reinforcement and a conditioned galvanic skin reflex with electric shock as UCS. The remaining 4 $Ss$ were experimentally naïve.

*Apparatus.*—Discrimination training and generalization testing were administered with $S$ seated in a primate chair enclosed within a light-tight cubicle (described elsewhere in greater detail, Ganz & Riesen, 1962). Essentially, the chair departs from current models in permitting control of head-position by the use of side-pieces above the neckboard. The neckboard is sufficiently wide to prevent the monkey from reaching his face. The $S$ wears light-diffusing contact lenses. Stimulus input is thus rendered largely independent of receptor-orienting behavior. There is a key underneath the neckboard. When a reinforcement is available, a key press will

activate a solenoid valve for 1 sec. This will, in turn, release 3–4 drops of 5% sucrose solution, directed to *S*'s mouth via tube; a 60-cycle buzzer connected in parallel with the solenoid provides secondary reinforcement.

The stimuli were produced by passing the beam of a 300-w. projection lamp through Bausch and Lomb second-order interference filters, a Wratten No. 8 filter, and plastic polarizing materials. The interference filters, Series 33–78 were used to obtain monochromatic pass-bands covering 449 through 631 m$\mu$ in nominal 10-m$\mu$ steps (range of step size is 4–18 m$\mu$; half band width averages 8 m$\mu$; peak transmission averages 35%). The yellow gelatin absorbed third-order interference peaks in the short wave lengths. The polarizing materials were used to approximate an equal luminosity spectrum. The energy transmission was estimated with a Weston photoelectric cell No. 8PV1AAF. By correcting for both the spectral sensitivity of the cell and of the human eye (photopic luminosity function) an estimate of the luminosity was obtained. Each interference filter was then coupled with a pair of suitably rotated polarizers to give an equal luminosity spectrum for the human *O*. The evidence suggests this is justified for the macaque, except possibly for the deep red region where luminosity could be lower for that *S* (Grether, 1939). The beam, after entering the cubicle, was finally focused by a lens system as a $\frac{3}{4}$-in. disc on *S*'s left eye; the right eye carried a black opaque contact lens. With the 564-m$\mu$ peak filter in place, illumination was estimated at 12.3 ft-c, using a Macbeth illuminometer in a heterochromatic match. The Tenite white contact lens diffused nonselectively in the visible region, with approximately 10% transmission.

The stimulus presentations were cycled automatically. White noise was delivered by a speaker within the cubicle to mask relay cues. Periodic checks revealed complete generalization across a variety of changes in auditory cues.

*Training sequence.*—During initial training, S$^d$ was continuously present for the 30-min. session. The response, a key press, was developed by a method of successive approximation with continuous reinforcement. Once criterion was attained—a minimum of 50 responses over 2 consecutive days—stimulus cycling was introduced with 15 sec. S$^d$ alternating with 5 sec. blackout and reinforcement was given on a 7.5-sec. variable interval schedule (VI), to a similar criterion level. Next, a 7.5-sec. delay of positive reinforcement (Ferster, 1958) was introduced to key

responses emitted during the blackout period, and again training was carried to the same criterion level. Next, the delay was increased to 15 sec. and carried to a criterion of a maximum of 20% responses during blackout over 2 consecutive days. Lastly, VI 15 sec. was introduced and training continued to a similar criterion (over 4 consecutive days). In the next sessions, generalization tests were administered. Following the tests, *S*s were introduced to discrimination training to wave length. The S$^d$ now alternated in irregular fashion with S$^\Delta$. During S$^d$, VI 7.5 sec. was reinstituted. In the presence of the blackout, or of S$^\Delta$, negative reinforcement was delivered by delaying positive reinforcement 7.5 sec. following each response (criterion: 80% responses emitted during S$^d$ and 100 responses per session, minimum over 2 consecutive days). Finally, the reinforcement and delay schedules were increased to 15 sec. Training was carried to a similar criterion, but over 4 days.

Generalization testing entailed the presentation of a series of seven hues, from S$^d$ through S$^\Delta$. The SGs were cycled in the same manner as the S$^d$ and S$^\Delta$, but without either type of reinforcement. There were 7 testing days and each followed the sequence: 10 training trials, four SG presentations, 13 training trials, three SG presentations, 30 training trials. For reasons which will be given in the discussion, it seemed advisable to keep the number of consecutive generalization stimuli as small as possible. The order of presentation of the seven SGs was randomized according to a 7 × 7 Latin square—Days × Order.

*Experimental design.*—The spectrum from 449 to 631 m$\mu$ was divided into three regions: 449–509 m$\mu$, 509–567 m$\mu$, and 567–631 m$\mu$. The ends of each of these regions comprised the S$^d$ and S$^\Delta$. This division achieves a rough symmetry, with a jnd maximum midway between 449 and 631 m$\mu$ (535 m$\mu$), a minimum 45 m$\mu$ from the ends (approximately 500 and 587 m$\mu$), and maxima at the ends of this spectrum (449 and 631 m$\mu$).

First, one of the hues 449, 509, 567, or 631 m$\mu$ was chosen as S$^d$ for the undiscriminated operant. Then, each *S* was trained to discriminate pairs of hues in all three regions in succession, with generalization testing following the attainment of each discrimination. The 6 *S*s were assigned to two groups with opposite S$^d$, S$^\Delta$ assignments. The successive discriminations did not involve any reversals in absolute stimulus value. Within each group, each *S* followed a different succession of discriminations and generalization tests,

## TABLE 1
### Training and Generalization Stimuli

| Response | Training | | | | Generalization | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Simple operant | 631 | 567 | 509 | or 449 | 631 | 604 | 567 | 540 | 509 | 483 | 449 |
| Discriminated operant | Group I | | Group II | | | | | | | | |
| | $S^d$ | $S^\Delta$ | $S^d$ | $S^\Delta$ | | | | | | | |
| | 631 | 567 | 567 | 631 | 631 | 625 | 610 | 604 | 588 | 580 | 567 |
| | 509 | 567 | 567 | 509 | 567 | 564 | 546 | 540 | 527 | 519 | 509 |
| | 509 | 449 | 449 | 509 | 509 | 503 | 489 | 483 | 469 | 457 | 449 |

Note.—Wave length in millimicrons.

as shown in Fig. 1. The details of this design are presented in Table 1.

### RESULTS

The experimental hypothesis postulates an inverse relationship between the steepness of generalization and the magnitude of jnd for the specific $S^d$. We predict, on the basis of the human differential hue function (Wright, 1947, Fig. 95), relative shallowness of gradient in the 630-, 530-, and 450-m$\mu$ regions (where jnd's are relatively large) and relative steepness in the 580- and 490-m$\mu$ regions (where jnd's are small).
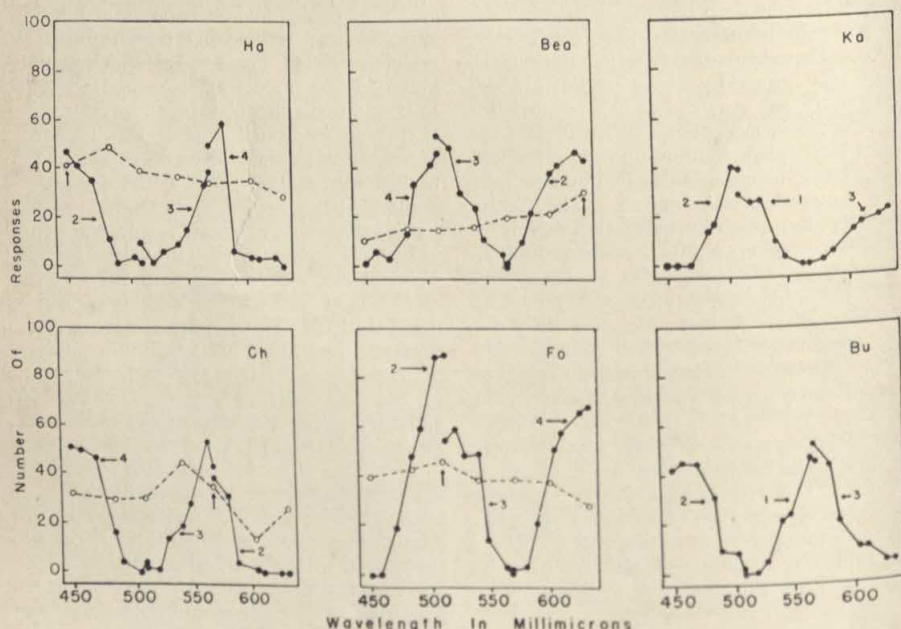


Fig. 1. Individual generalization gradients. (The dashed lines designate the simple operant gradients, arrow at the $S^d$; the solid lines designate the discriminated operant gradients. The numbers identify the order of administration of the successive discriminations and generalizations. Six Ss are depicted here.)

Gradients were obtained by simply summing responses obtained by each SG. Figure 1 depicts the generalization of both the undiscriminated and the three discriminated operants of the 6 Ss. Generalization of the undiscriminated operant was measured in only 4 Ss. It is apparent, first, that the undiscriminated operant gave an almost horizontal gradient. There is a mild downward tendency which does not appear related to jnd magnitude in the manner predicted. Such near-horizontal gradients are of little value in testing the effect of differences in jnd size. Any differences cancel out when cumulated over more than 50 mμ of wave length.

The three gradients following discrimination training are much steeper. Their shapes are, in almost all cases, bell-shaped, i.e., the response decrement first accelerates and then decelerates. From a casual perusal, it

appears Ss Ha, Ch, and Bu show steep gradients (descending) in the 580-mμ region (jnd small); Ss Ka, Fo, and Be show steepness in the 490-mμ region (jnd small); Ha, Ch, and Bu show shallowness of gradient in the 450-mμ region (jnd large); Ka and Be show shallowness of gradient in the 630-mμ region (jnd large). These gradient differences are in the direction predicted by the experimental hypothesis.

If we take each gradient and express the number of responses emitted during an SG presentation as a percentage of the maximum emitted during any of the seven SGs, we can plot relative gradients, which make slope comparisons easier. This is done for all 6 Ss in Fig. 2. The following would appear consistent for all 3 Ss in Group I: the 509–449 mμ gradient is steeper than the 631–567 mμ gradient in its descent from S^d to
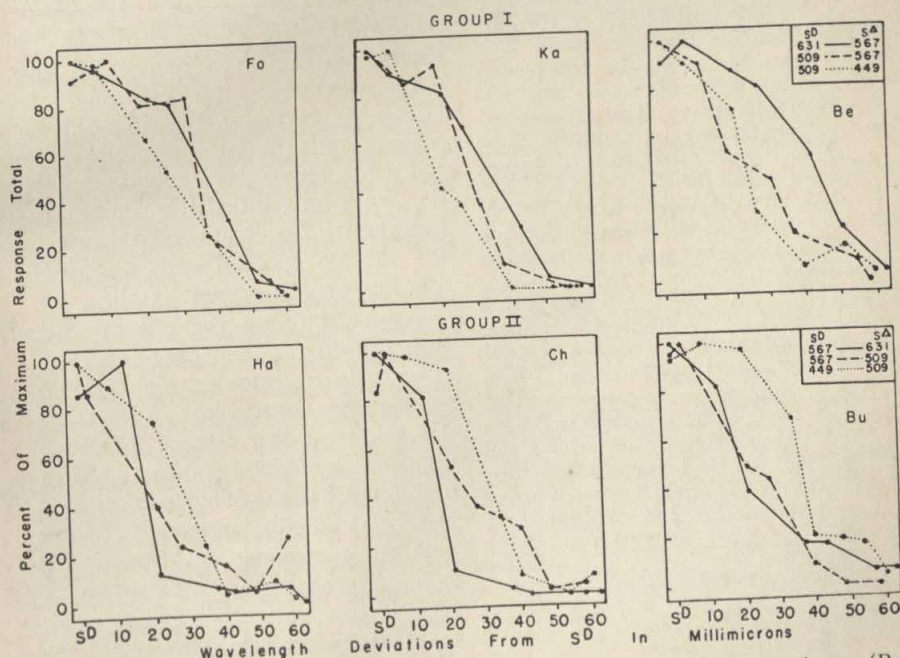


FIG. 2. Individual generalization gradients of 6 Ss in three adjoining spectral regions. (Response totals are here expressed as percentages of the maximum response total.)

$S^\Delta$ (in this case the 631-m$\mu$ $S^d$ has a larger jnd than the 509-m$\mu$ $S^d$). For Group II the 3 $S$s show shallowness in the slope of descent in the 449–509 m$\mu$ region (jnd at the 449-m$\mu$ $S^d$ is relatively large) and $S$s Ha and Ch show delayed ascent in the 631-m$\mu$ region (jnd at the 631-m$\mu$ $S^\Delta$ is relatively large). Moreover, the relationships between the three gradients in Group I are reversed in Group II. This should follow if jnd's are exercising any effect since the $S^d$, $S^\Delta$ assignments are reversed from Group I to Group II. The results suggest a generalization-discriminability relationship in the direction predicted. We now turn to a statistical evaluation.

As is customarily the case with stimulus generalization gradients, response strength was highly correlated with its variability, thus precluding analysis of variance either on the raw data or on any customary transformations. Our main interest centered on some simple index of gradient slope. The slope of a generalization curve can be viewed as a measure of the variability of the stimuli which can elicit the training response. This can be estimated by taking the cross product of each $S^d$-SG distance in millimicrons and the number of responses that SG elicited, summing over the seven cross products of a test day, and dividing by the total number of responses elicited that day. We obtain an average deviation statistic which is directly proportional to slope shallowness. Analysis of variance, mixed design, was performed on this statistic. If changes in jnd size modify gradient slope, this would be reflected here as a significant Group × Region interaction, which was in fact the case with a $P < .02$ ($F = 7.45$; $df = 2/24$). The design, it will be recalled, placed one group's predicted shallowness of gradient against the other's predicted steepness, thereby achieving a null combined effect for the two groups together. Accordingly, the Regions SS was not above chance expectation ($F < 1.00$). A significant Day SS ($F = 2.89$; $df = 6/8$; $P < .05$) reflects the effect of extinction, the gradients becoming steeper on successive testing days.

Are the magnitudes of these gradient differences in proportion to the magnitudes of the corresponding jnd

differences? Since the peak transmissions of the filters are not spaced along equal steps, gradients across regions can be compared only if one interpolates, linearly in this case, along 10-m$\mu$ steps. It is assumed that for distances of 10 m$\mu$, rhesus generalization gradients do not depart seriously from linearity, which seems not unreasonable at the present level of precision. The question of interest is the departure of some specific gradient from an average gradient, i.e., with a difference gradient. Therefore, the three gradients of each $S$ were combined to obtain his average gradient, and then the individual gradients were expressed as deviations from this average. These are shown in Fig. 3 for the individual $S$s. Values above zero reflect relatively shallow gradients; negative values reflect relatively steep gradients. Each of the six graphs represents a group's performance in one region. It can be seen, first, that a certain measure of inter-$S$ agreement is present in gradient deviations, for example, the Group II curves in the 567–631 m$\mu$ region, and the Group I curves for the same region. This again reflects the wave-length-tied discriminability function. Systematic deviations for the middle spectral region, 567–509 m$\mu$, are weak if present at all.

Figure 3 brings to light an interesting finding. There are never deviations from average gradients at the $S^d$ and $S^\Delta$ themselves. The largest deviations appear *between* $S^d$ and $S^\Delta$. This is interesting because at 631 m$\mu$, e.g., one of the training stimuli, there is a jnd decidedly larger than average, yet no corresponding deviation in generalization; the same can be said for 449 m$\mu$. At other loci, e.g., 480 and 600 m$\mu$, positioned between $S^d$ and $S^\Delta$, there is a strong deviation from average both in the magnitude of the
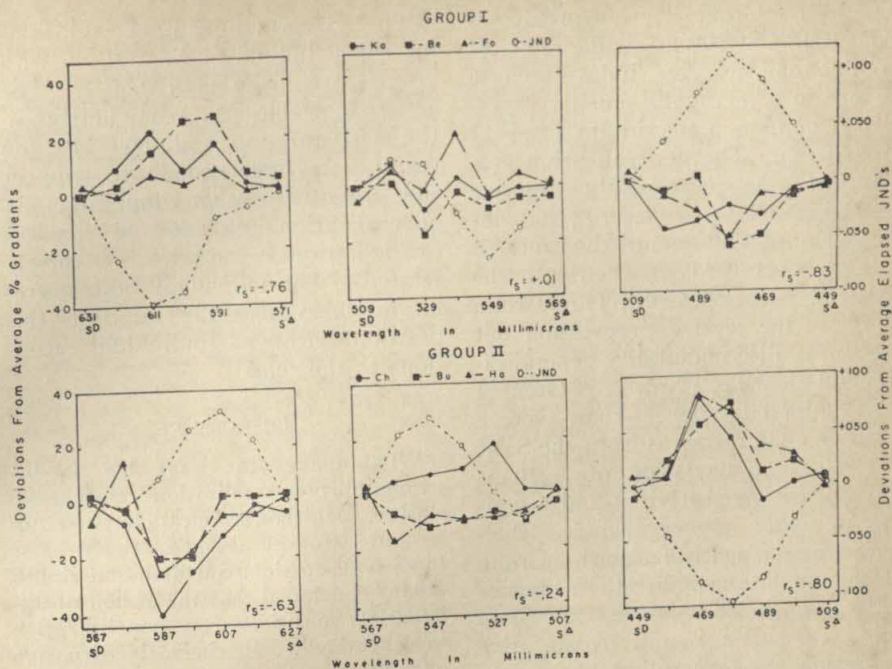
FIG. 3.  Deviation gradients in three spectral regions.  (Each figure depicts one group's performance in one region.  The filled data points record deviations of individual subjects from their own average gradients.  The dashed line depicts deviations from the cumulated average jnd.)

jnd and in the generalization gradient. This seems to imply that the generalization gradient is responsive, in the present experiment, to the deviations in the cumulated jnd, not to deviations in the jnd itself.

A measure of deviations in cumulated jnd's was derived.  For the $i$th spectral region, $(i = 449$–$509$ m$\mu$; $509$–$567$ m$\mu$; $567$–$631$ m$\mu$), an average jnd was first computed.  The average jnd for the region $i$ equals the difference between $S_i^d$ and $S_i^\Delta$ in millimicrons divided by the number of jnd's included between $S_i^d$ and $S_i^\Delta$. A specific generalization stimulus in the region $i$ is then taken, $SG_{ij}$ ($j = 0$, $10$ m$\mu$, $20$ m$\mu$ . . . etc. from $S^d$) and the distance is counted off between the training stimulus $S_i^d$ and $SG_{ij}$ in average jnd's; also that same distance

is counted off in the jnd's obtained from the human $\Delta\lambda$ function.  The difference between the two counts gives the measure desired, depicted by the dashed lines in Fig. 3.  Negative values reflect an accumulation of larger-than-average jnd's (poor sensitivity); positive values reflect an accumulation of smaller-than-average jnd's (high sensitivity).  By definition, this measure is zero both at $S^d$ and $S^\Delta$.  Figure 3 shows the function to be, if anything, inversely related to the gradient deviations. This is clearest in the 449–509 m$\mu$ region where Group II has, from 449 to 489 m$\mu$ accumulated larger-than-average jnd's, thus giving a deviation of $-.11$ at 479 m$\mu$ (poor sensitivity) and positive deviation in generalization gradient at that point (shallow

slope). For Group I, as one progresses from the $S^d$ at 509 m$\mu$ one cumulates smaller-than-average jnd's because we are in a relatively sensitive spectral region, to a maximum of $+.11$ at 479 m$\mu$. The generalization gradient has an increasingly negative deviation (slope is steep) that increases until 479 m$\mu$ and then returns to zero when $S^\Delta$ is reached. In the 631–567 m$\mu$ region, Group I and Group II show the reverse progression but the same discriminability-generalization relationship. Now it is Group I that cumulates larger-than-average jnd's as one progresses from the $S^d$ at 631 m$\mu$ to a deviation of $-.10$ at 611 m$\mu$ (poor sensitivity) and a positive deviation in generalization (shallow slope of generalization). Group II accumulates smaller-than-average jnd's from the $S^d$ at 567 m$\mu$ to 607 m$\mu$ (high sensitivity) and a negative deviation in generalization (steep slope of generalization). In the middle region, 567–509 m$\mu$, where the cumulated-jnd deviation function reverses in midstream, the relationship is either absent or so attenuated as to be indiscernible.

TABLE 2

CORRELATIONS BETWEEN GENERALIZATION
GRADIENT DEVIATIONS AND THE
CUMULATED jnd DEVIATION
FUNCTION

| Region | Subjects | | | |
|---|---|---|---|---|
| Group I: | Ka | Be | Fo | Comb. |
| 631–567 m$\mu$ | $-.83*$ | $-.47$ | $.08$ | $-.76*$ |
| 509–567 m$\mu$ | $.35$ | $.40$ | $.19$ | $.01$ |
| 509–449 m$\mu$ | $-.51$ | $-.51$ | $-.76*$ | $-.83*$ |
| Group II: | Ch | Bu | Ha | Comb. |
| 567–631 m$\mu$ | $-.38$ | $-.19$ | $-.47$ | $-.63*$ |
| 567–509 m$\mu$ | $.56$ | $-.40$ | $-.65*$ | $-.24$ |
| 449–509 m$\mu$ | $-.44$ | $-.87*$ | $-.81*$ | $-.80*$ |

$* P < .05.$

The Spearman rank correlations for the 6 individual $S$s and the two groups are given in Table 2. The group correlations, particularly in the 631–567 m$\mu$ and 509–449 m$\mu$ regions are large enough to support the hypothesis that deviations in generalization gradient and *cumulative* deviations in jnd rate are inversely related. The individual rank correlations also generally support this position, with some individual exceptions to the rule.

## DISCUSSION

What makes the wave length discriminability-generalization relationship elusive, such that it appears in two studies and is totally absent in two others? In their discussion, Guttman and Kalish (1956) concluded that the action of differential sensitivity on generalization was effective in a discriminative situation but not in the emission of a simple operant. We will now continue this discussion in the light of subsequent studies.

The first reason we propose why this is so is that without discrimination training, generalization gradients include more error variance. This in turn masks the rather subtle differential effect of the jnd. This additional error variance arises from at least two sources. One has to do with the inevitable discrimination training which occurs during the development of a simple operant. For example, this occurs at a moment when the pigeon is looking at the side of the box and pecking the wall, instead of the key. Such discriminations necessarily steepen gradients. Moreover, this adventitious discrimination training will necessarily vary from $S$ to $S$ and will set up inter-$S$ variations in gradients. If within-$S$ gradient comparisons are made, error variance arises from the course of this adventitious discrimination training at the successive $S^d$. The $E$'s introduction of discrimination training to a percentage criterion level has the effect of making such training explicit and uniform for

the stimulus dimension tested. It anchors the slope of the gradient at two points on the continuum, $S^d$ and $S^\Delta$. A second source of variability arises from stimulus preferences. For example, it is well established that the pigeon has a proclivity for 580 m$\mu$. When tested for generalization, it appears to generalize strongly into 580 m$\mu$. This is puzzling sometimes because at 580 m$\mu$ the jnd is quite small (high sensitivity). It appears likely that there is merely more pretraining response strength at 580 m$\mu$. In discrimination training such preferences are weakened. They can be isolated from jnd effects by balancing $S^d$, $S^\Delta$ assignments in two groups.

The second and more fundamental reason why the jnd effect on discrimination is clear in a discrimination experiment and not in the generalization of a simple operant is that in a discrimination experiment we usually present only two stimuli at some arbitrary interval while in a generalization experiment we customarily present a repeated series at equal physical intervals. Repeated series of generalization trials were given in the Guttman and Kalish (1956) experiment (two series of 132 extinction trials) and in Blough's (1961) study (three initial series of 66 trials). The generalization gradient is probably not invariant across changes in the distribution of the generalization stimuli: the "frame of reference" effect (Humphreys, 1939). When SGs are densely distributed, there will be less generalization. This follows in part from the fact that extinction also generalizes and summates. Suppose, as in previous studies, that the SGs are in uniform steps of 10 m$\mu$, but with jnd's of varied sizes. If a series of GSs are presented repetitively under extinction, extinction will cumulate and generalize from one GS to another. Where jnd's are large, a shallow gradient of generalization is expected. Thus more positive response strength will generalize from $S^d$ to GS. But more negative response strength resulting from extinction will also generalize, from one GS to all the others. If this is the case, extinction effects generated by equally spaced GSs

would act precisely in such a manner as to cancel the positive differential effects of the jnd. The final result is that $S$ appears to be generalizing to physical wave-length distances, rather than to the differential properties of his receptor system. The result is dependent on $E$'s choice of GS spaced at physically equal 10-m$\mu$ steps. In discrimination training, since only two stimuli are presented during each training sequence, the same cancellation effect should not appear and the jnd is then seen to be effective.

The balancing of positive and negative response strength to cancel out the jnd effect was not manifested in the Kalish (1958) study because there is no extinction, in the usual meaning of the term, during psychophysical measurements on human $S$s, e.g., using the method of single stimuli. In the present study, this cancellation was not realized, first, because the stimuli were not evenly spaced, and second, because no series longer than four was presented. Thus, for different reasons, both studies yielded positive results.

## SUMMARY

An experiment evaluated the widely held view that slope of generalization gradient and resolution capacity are inversely related. Four rhesus macaques were trained to emit a simple operant to a monochromatic hue in the 450–630 m$\mu$ range. Generalization was then measured under extinction across this range. This gradient was, in all cases, almost horizontal. Operants discriminated with respect to wave length ($S^d - S^\Delta$ about 60 m$\mu$ apart) were then developed in these $S$s and 2 additional $S$s. Generalization was measured as before but in nominal 10-m$\mu$ steps. Each $S$ was trained to discriminate and generalize three successive pairs of stimuli. A number of the gradients revealed differences in slope which were in accordance with the predicted inverse relationship. Analysis of the deviations in gradient slope suggested these were related not to deviations from the average jnd size but, rather, to deviations from average cumulated jnd's. A measure of deviations from average cumulated jnd's was derived and a high negative correlation to the gradient deviations was shown to exist. It was concluded that some infrahuman generaliza-

tion gradients do reflect the $S^d$-SG distance in cumulated jnd's. Some factors that may have masked this relationship previously were discussed.

## References

BLOUGH, D. S. The shape of some wavelength generalization gradients. *J. exp. Anal. Behav.*, 1961, **4**, 31–40.

BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

FERSTER, C. B. Control of behavior of chimpanzees and pigeons by time out from positive reinforcement. *Psychol. Monogr.*, 1958, **72** (8, Whole No. 461).

GANZ, L., & RIESEN, A. H. Stimulus generalization to hue in the dark-reared macaque. *J. comp. physiol. Psychol.*, 1962, **55**, 92–99.

GEWIRTZ, J. L., JONES, L. V., & WAERNERYD, K. Stimulus units and range of experienced stimuli as determinants of generalization-discrimination gradients. *J. exp. Psychol.*, 1956, **52**, 51–57.

GRETHER, W. F. Color vision and color blindness in monkeys. *Comp. psychol. Monogr.*, 1939, **15** (4, Whole No. 76).

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HUMPHREYS, L. G. Generalization as a function of the method of reinforcement. *J. exp. Psychol.*, 1939, **25**, 361–372.

KALISH, H. I. The relationship between discriminability and generalization: A re-evaluation. *J. exp. Psychol.*, 1958, **55**, 637–644.

LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

WRIGHT, W. D. *Researches on normal and defective colour vision.* St. Louis: C. V. Mosby, 1947.

# EFFECT OF PATTERN AND PLEONASM LOCATION IN SERIAL LISTS UPON ACQUISITION AND SERIAL POSITION ERRORS [1]

RONALD L. ERNST,[2] CHARLES P. THOMPSON, AND W. J. BROGDEN

*University of Wisconsin*

In a series of studies on verbal maze learning (Ernst, Hoffeld, Seidenstein, & Brogden, 1960; Namikas, Thompson, & Brogden, 1960; Thompson, Voss, & Brogden, 1957), the form of the serial position error curve was found to be altered by the location of a pleonasm in the standard maze pattern. The pleonasms used were a doublet (two successive identical items), a split-doublet (two identical items with a different item between them), a triplet (three successive identical items), and a quadruplicate (four successive identical items). Although there is no evidence that any of the experimental patterns produce any difference in acquisition from that for the standard pattern, each produces a significant, characteristic difference in the form of the serial position error curve at the locus of the pleonasm. Locus of pleonasm is significant for certain experimental patterns and not for others. The pleonasm effect may vary as a function of stage of practice and of the error measure, whether all errors per trial or only the first response per trial (first errors) are considered.

Verbal maze learning necessarily involves the correction procedure and a discovery phase that may be distinct from and a prerequisite to a later acquisition phase (Melton,

1950). If the same list represented by a maze is learned by the serial anticipation method, the noncorrection procedure is a necessary component and the discovery phase is either absent or of much lesser magnitude. These differences between methods suggest that the effects of patterns including pleonasms will be different. Because the first error measure of maze learning is comparable to the error measure of serial anticipation learning and because there is little or no discovery phase with serial anticipation learning, the effects of pleonasms for this procedure are likely to be represented by decreases in error relative to comparable positions for the control list, and acquisition of the experimental patterns should require fewer trials and show fewer total errors than the control list. The experiment to be reported was designed to test the above hypotheses.

## METHOD

*Design.*—The experimental design consists of four patterns, each with a single pleonasm (doublet, split-doublet, triplet, or quadruplicate) at three locations (early, middle, or late in the list) and the control list composed of the numbers 10, 20, 30, and 40, each occurring four times to provide a length of 16. The control list is identical with the control maze pattern used in previous studies (Ernst et al., 1960; Namikas et al., 1960; Thompson et al., 1957). Each experimental pattern is a limited modification of the control pattern that provides for the location of the appropriate pleonasm without changing the frequency of the four numbers, introducing other pleonasms elsewhere, or altering the length of the list. Each of the 12 experimental patterns was

[2] Now at Carnegie Institute of Technology.

permuted to obtain the four sublists so that each of the four numbers (10, 20, 30, 40) occurs once in each of the 16 positions.

*Subjects.*—The *S*s were 312 University of Wisconsin students of elementary psychology. There were 24 *S*s in each of the 13 groups, and 6 *S*s in each group learned one of the four subpatterns. Each of 2 *E*s tested one-half of the *S*s for each subpattern of each group. Assignment of *S*s to *E* and to experimental conditions was random within each replication.

*Procedure.*—The instructions of Thompson, Voss, and Brogden (1957) were modified as follows to make them appropriate to serial anticipation learning and were read by *E* to each *S*.

In this experiment you will be required to learn a single list of numbers. These are the numbers used: 10, 20, 30, and 40. Each number will be used more than once in the list. The first trial is a study trial and will show you the series of numbers. On the next and all subsequent trials you will try to anticipate correctly each successive number in sequence. For example, when the word START is presented on the screen, you will try to respond with the first number of the series. When the first number is on the screen, you will try to respond with the next number, and so on to the end of the trial. In other words, you will try to keep one step ahead of the projector making only one response per exposure. Please respond as fast as you can. When the first trial is completed, there will be a short rest before the second trial is started. We will continue with similar trials until you are able to make one repetition of the series without error. Are there any questions?

After answering any questions, the appropriate list was presented to *S* by means of a Dunning Animatic strip film projector. The duration of each stimulus item was 2 sec. and the items followed each other in sequence until the end of the list was reached. The intertrial interval was 30 sec. The projection screen was a 5.5 × 7.5 in. rectangle of frosted glass set in a 2 × 3 ft. piece of plywood painted flat black. The screen and projector were placed on a table at the screen end of which *S* was seated with an approximate viewing distance of 30 in. The stimuli were ¾ in. high at the screen. The procedure described in the instructions was maintained until *S* reached the criterion of one errorless trial. The *E* recorded the response for each position of the list on each trial.

## RESULTS

Evaluation of the effects of pattern and pleonasm location in the serial list upon speed of acquisition was made by separate analyses of variance each for trials and errors to the criterion for each pattern and its pleonasm location versus the control list, and between all patterns and pleonasm location excluding the data for the control list. Tests of homogeneity of variance show the data to be heterogeneous. Because of this, statistical significance for these and all subsequent analyses was set at the 1% level in lieu of the 5% level. Of the separate analyses, the source of variation represented by location of pleonasm versus the control pattern show significant *F* ratios for the split-doublet and triplet with the trial measure and for the split-doublet, triplet, and quadruplicate with the error measure. In the analyses between all patterns and pleonasm location with the control data excluded, pattern, pleonasm locus, and the interaction of pattern and pleonasm locus are significant for both trials and errors. No other source of variation is significant in any of the above analyses except that for subpattern (permutation of numbers) in the case of the doublet with the trial measure, and this is without meaning in the experimental design.

The means represented in the analyses are presented in Table 1. Range tests (Duncan, 1951) of the differences between means for significant sources of variation in the analyses of variance give almost identical results for the trial and error data. For the split-doublet, the mean for Locus 1 (early) is significantly smaller than the means for Locus 2 (middle), Locus 3 (late), and the control pattern, which do not differ significantly among them-

TABLE 1

GROUP MEAN TRIALS AND ERRORS TO ACQUISITION CRITERION

| Pleonasm | Pleonasm Locus | | | | | | Overall Mean | |
| | 1 (Early) | | 2 (Middle) | | 3 (Late) | | | |
| | Trials | Errors | Trials | Errors | Trials | Errors | Trials | Errors |
| Doublet | 11.83 | 66.75 | 12.75 | 68.08 | 11.38 | 61.92 | 11.99 | 65.58 |
| Split-doublet | 8.33 | 47.58 | 18.42 | 111.13 | 19.71 | 110.17 | 15.65 | 89.63 |
| Triplet | 9.00 | 40.79 | 9.04 | 40.96 | 7.88 | 42.83 | 8.64 | 41.53 |
| Quadruplicate | 10.17 | 44.88 | 9.67 | 44.42 | 11.75 | 59.79 | 10.53 | 49.69 |
| Overall mean | 9.96 | 50.00 | 12.45 | 66.15 | 12.68 | 68.68 | | |
| Control group | Mean Trials = 14.88 | | Mean Errors = 81.33 | | | | | |

selves. All means of the triplicate are significantly smaller than that for the control, but are not significantly different from each other. All means for the quadruplicate (error measure only) are significantly smaller than that for the control list, but do not differ among themselves. Of the overall means for locus regardless of pattern, that for Locus 1 is significantly smaller than those for Loci 2 and 3. Of the overall means for pattern regardless of locus, the mean for the split-doublet is significantly larger than those for all other patterns, and the mean for the triplet is significantly smaller than that for the doublet. Of the means representing the interaction of pattern and locus of ploenasm, those for Loci 2 and 3 for the split-doublet are significantly larger than all other means. There are no other significant differences between the means.

Initial analysis of the form of the serial position error curves was accomplished by a separate analysis of variance for each kind of pleonasm, involving the data for the three loci of pleonasm and the control list. In each of these four analyses, significant $F$ ratios were obtained for serial position, and for the interaction of serial position and pleonasm locus. Since there are significant differences in total errors to the acquisition criterion for pleonasm location in the split-doublet pattern, and between the control and other patterns, the error data for each $S$ at each serial position were converted to percentage of his total errors. Analyses of variance completed on the transformed data also show uniform results of significant $F$ ratios for serial position and the interaction of serial position and pleonasm locus. The differences in the form of the serial position error curves are shown in Fig. 1 where each experimental curve is presented in comparison with the curve for the control pattern. Although most of the experimental curves show differences in form from the control curve at the positions of the pleonasm, there also are differences between the experimental curves and the control curve at other serial positions. Since these latter differences do not appear to bear any relationship to any variables of pattern, all further analyses of the effect of pattern upon the form of the serial position error curve are restricted to those positions at which pleonasms are located.

The difference in percentage of
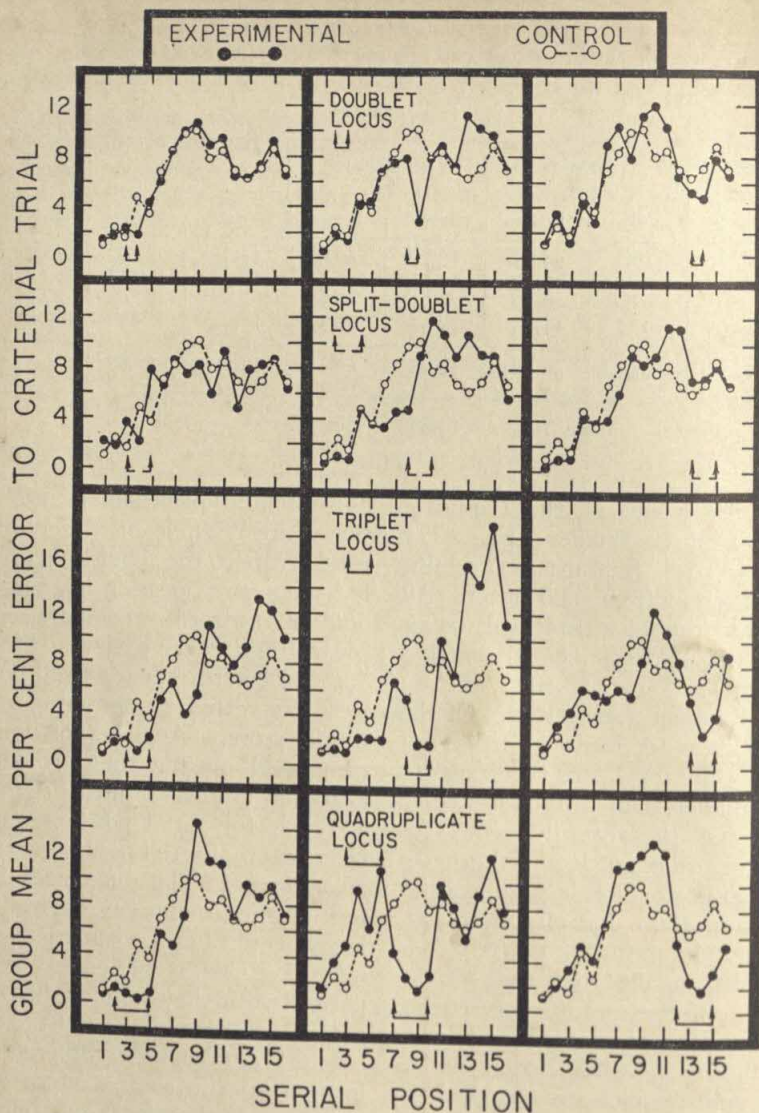
Fig. 1. Serial position error curves in terms of percentage of total errors as a function of pleonasm and pleonasm location.

total error for each position of the pleonasm (plus the preceding and following position) from the mean percentage of total error for the control group at the identical serial position was computed. Analyses of variance of the data for the three pleonasm locations for each of the four pleonasm types were completed. In each of these four analyses serial position and the interaction of serial position and pleonasm locus are significant sources of variation. The $t$ test was used to establish fiducial limits for each of the four sets of data at the 1% level for the means over loci thus providing for assessment of the serial position effects shown in the far right column of Fig. 2 and for the means of the three loci, thus providing for assessment of the interaction of serial position and location

shown for each row in the three left hand columns of Fig. 2. For the data over all loci, the effect for the doublet is a decrease in error at the second position; the split-doublet is an increase in error at the third position; the triplet is a decrease in error at the second and third positions; and the quadruplicate is a decrease in error at all four positions. Range tests (Duncan, 1951) show no significant differences in magnitude of decrease in error as a function of position for either the triplet or the quadruplicate.

Since the detailed results of the analyses of the interaction of pleonasm and locus are complex, only the general results are noted. The maximum effect for all pleonasms is at Locus 2. The minimum effect occurs at Locus 1 for the triplet and quadruplicate and at Locus 3 for the doublet and split-doublet. There is a unique form of the split-doublet at each locus, but the other pleonasms show consistency in form over loci.

The error data were tabulated for each $S$ for each half of trials to the criterion for each of the positions of the



Fig. 2. Pleonasm effects as a function of locus within the series and position within the pleonasm. (The dotted horizontal lines represent the fiducial limits at the 1% level of confidence.)

pleonasms plus the preceding and following position, and converted to percentage of total error. The difference from the mean of similar measures for the control list was obtained and analyses of variance were conducted separately for each pleonasm. Stage of learning is a significant source of variation for the triplet, for the interaction of stage of learning and pleonasm locus for the triplet and quadruplicate, and for the triple interaction of stage of learning, pleonasm locus, and serial position for the split-doublet, triplet, and quadruplicate. All of these effects indicate fewer errors during the second half of trials than during the first half except that the triple interaction for the split-doublet indicates an increase in error for the second half of trials at the third position for Locus 2. The form of the curves for the first and second halves of trials is substantially the same in all cases except for the split-doublet at Locus 2.

## DISCUSSION

The hypotheses which the experiment was designed to test were confirmed in part. The introduction of pleonasms either increases speed of acquisition over that for the control list or has no effect. Increased speed of acquisition occurs for the triplet and quadruplicate and for one locus of the split-doublet. No differences in speed of acquisition occurs for the doublet and for two loci of the split-doublet. Pleonasm locus has a differential effect upon the form of the serial position error curve in all cases. Maximum effect occurs for pleonasm locus in the middle of the list. The least effect occurs for the triplet and quadruplicate with early locus and for the doublet and split-doublet with late locus. The nature of the effect for a given pleonasm is similar at all loci, except for the split-doublet. The mean effect over all loci is a decrease in error at the second position for the doublet; an increase in error at the third position for the split-doublet; a decrease in error at Positions 2 and 3
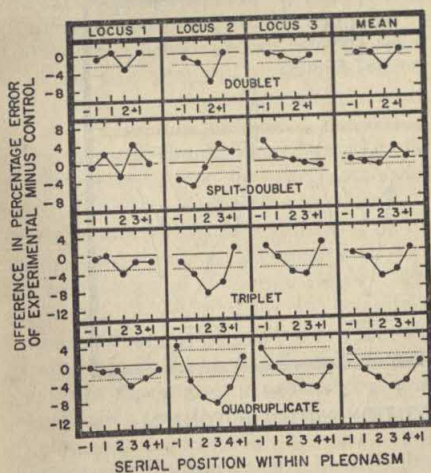
for the triplet; an increase in error at the −1 position and decreases in error at Positions 1, 2, 3, and 4 for the quadruplicate. Stage of learning is not a significant factor in the form of the pleonasm effects except in the case of the split-doublet.

In contrasting maze and serial anticipation learning, the control pattern or list is acquired with a mean time of 772.9 sec. and a mean of 110.8 total first errors as a maze (Ernst et al., 1960) and with a mean time of 476.2 sec. and a mean of 81.3 total errors as a serial list. Both differences in favor of serial learning are statistically significant. These results are comparable to those obtained by Thompson and Brogden (1958) in comparing the correction procedure with a modified correction procedure similar to the procedure for serial anticipation learning. Serial position error curves for the control list in percentage of total error are presented in Fig. 3 for maze and serial learning. The form of these curves appears to be substantially the same for the two methods of acquisition and because of this no statistical analysis of difference in form was made.

The effect of locus of pleonasm for maze and serial learning cannot be com-
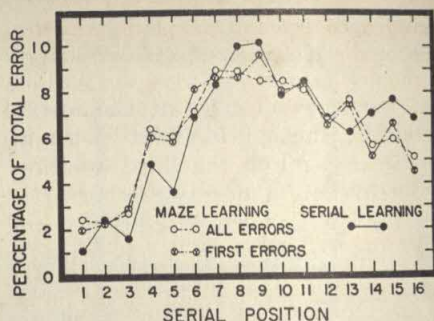


FIG. 3. Serial position error curves for the control list acquired by maze and serial anticipation learning. (The all-error and first-error curves for maze learning are from data of Ernst et al., 1960.)

pared precisely because different locations were used in the two sets of experiments. Comparison of pleonasm effect, regardless of locus, is possible but only in terms of the first error measure of maze learning and the error measure of serial learning (Thompson & Brogden, 1958). Figure 4 presents curves of percentage of total error for the appropriate pattern positions. Curves for the all-error measure of maze learning are also
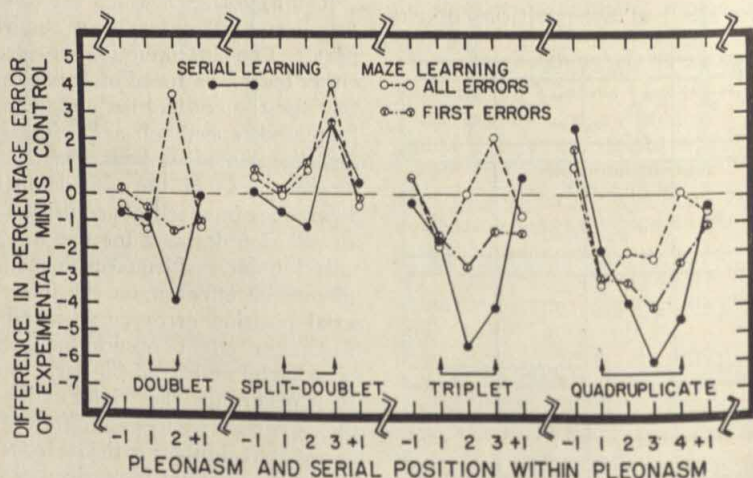


FIG. 4. Pleonasm effect as a function of serial anticipation learning and of the all-error and first-error measures of maze learning. (The curves of maze learning for the doublet and split-doublet pleonasms are from data of Ernst et al., 1960, and those for the triplet and quadruplicate are from data of Namikas et al., 1960.)

included to provide complete presentation of the pleonasm effects. The form of the first error curves for maze learning and the curves for serial learning are strikingly similar. The relative magnitude of the effect is greater for serial learning than for maze learning for the doublet, triplet, and quadruplicate. This difference in magnitude is probably due to the discovery phase inherent in the maze method which results in an initial inhibitory effect of the pleonasm, and differences in form as a function of stage of practice.

The split-doublet effect is similar for serial learning and both the first-error and all-error measures of maze learning. It is noteworthy in comparing the split-doublet effect to those for other pleonasms that it consistently occurs as an increment in error at the third position. The other pleonasms uniformly produce decrements in error except for the all-error measure for maze learning in the case of the doublet and triplet, and the position preceding the first position of the quadruplicate by all procedures.

In considering the learning of lists of items comparable to digits by the maze and serial anticipation methods, a more rapid acquisition in terms of both time and errors should always occur for the latter method. The effects of pure pleonasms such as doublets, triplets, and quadruplicates should be reductions in error for learning by noncorrection and for first-error measures of learning by the correction method. The reductions in error should be greater for the noncorrection than for the correction method and of sufficient magnitude to produce more efficient acquisition of the series containing a pleonasm than a random series of the same items. The discovery phase inherent in correction learning is responsible for the difference in magnitude just noted and also for the error increments produced by pleonasms when the all-error measure is used. Pleonasms represented by the split-doublet should have a consistent effect of error increment regardless of correction or noncorrection learning, stage of practice, or kind of error measure.

## SUMMARY

An experiment tested the effect upon acquisition and the form of the serial position error curve of type of pleonasm and pleonasm locus. The $Ss$ learned by serial anticipation a series of 16 numbers, composed of 10, 20, 30, and 40 each occurring four times. The control list was a random sequence without pleonasms. The experimental lists included one pleonasm (doublet, split-doublet, triplet, or quadruplicate) at one of three loci (early, middle, or late in the list). Acquisition was consistently faster for lists with a triplet or quadruplicate than for the control list. The form of the serial position error curve was altered by each pleonasm and there was significant interaction in form of the curve and locus of pleonasm. The doublet effect is a decrease of error at the second position, the split-doublet effect is an increase of error at the third position, the triplet effect is a decrease of error at the last two positions, and for the quadruplicate is an increase of error at the position just preceding and a decrease of error for all four positions of the pleonasm. Comparison of the effects of these pleonasms obtained in earlier experiments on maze learning was made with the results of the present experiment.

## REFERENCES

DUNCAN, R. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, **2**, 171–189.

ERNST, R. L., HOFFELD, D. R., SEIDENSTEIN, S., & BROGDEN, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (Rev. ed.) New York: Macmillan, 1950. Pp. 668–690.

NAMIKAS, G., THOMPSON, C. P., & BROGDEN, W. J. Effect of triplet and quadruplicate location in verbal maze patterns upon serial position errors. *J. exp. Psychol.*, 1960, **59**, 383–390.

THOMPSON, R. F., & BROGDEN, W. J. Acquisition of a verbal maze as a function of method of correction and number of alternate choices per unit. *J. exp. Psychol.*, 1958, **56**, 501–506.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

# A TEST OF THE ALL-OR-NONE HYPOTHESIS FOR VERBAL LEARNING [1]

## JOANNA P. WILLIAMS [2]

*Yale University*

The view that associations develop gradually with repeated pairings of the stimulus and response has been a basic tenet of most theories of learning (Hull, 1943; Spence, 1956; Thorndike, 1932). Recently the validity of this assumption has been questioned (Estes, 1960; Rock, 1957); it has been suggested instead that in learning situations involving simple stimuli and responses, acquisition occurs on an all-or-none basis. Repetition facilitates learning in such cases simply by providing a greater number of opportunities for an association to be formed. A core assumption of this position is that there are no "strengthening" effects of trials previous to the one on which the association is formed; therefore a response will have the same probability of being learned on each successive trial. From his examination of data from the first two trials of a paired-associate experiment, analyzed in such a way as to remove artifacts arising from averaging, Estes (1960; Estes, Hopkins, & Crothers, 1960) concluded that the probability of recalling an item on a given trial for the first time was in fact constant over trials.

It should be noted that this finding is not inconsistent with a strength theory, such as that of Hull (1943) or Spence (1956). According to such

a theory, associative strength must exceed the threshold of recall before a correct response is made. The easier the item, the fewer the repetitions required to bring it to threshold. Thus, on the first trials of a paired-associate list, the easier items will be recalled; on later trials, the more difficult items, having had the benefit of several repetitions, can reach threshold. In such a manner, according to the strength position, either an increase, a decrease, or no change in the mean probability of initial recall may be produced over trials by the proper selection of items of various degrees of difficulty. For an all-or-none approach, however, item heterogeneity could explain a decrease in probability over trials (easy items are quickly learned and are removed from the set of unlearned items so that more difficult items are selected as trials proceed), but it could not account for an increase.

Estes has limited his model to the one response measure of recall. However, an all-or-none position might be expected to have implications for other commonly used response measures as well. For example, it is likely that response latency would be sensitive to possible changes in response strength, especially in later stages of learning, when recall is nearly 100%. Latency has been used in previous studies (e.g., Brown & Huda, 1961; Simley, 1933) but has not been analyzed in a way relevant to the all-or-none hypothesis.

The present experiment was designed, using simple paired-associate

learning materials, (a) to test one of the fundamental assumptions of the all-or-none model—that the mean probability of initial recall is constant —over an extended number of trials, and (b) to examine the all-or-none position in the light of a continuous measure of learning, response latency.

In addition, the degree to which items in such a simple paired-associate situation can be analyzed as independent units was investigated. All-or-none data might be produced if S paid attention to and rehearsed only a few items on each presentation of the list; this kind of behavior, of course, would make a "trial" non-equivalent for the various items. For example, once an item is learned, the one which precedes it in the list might have a higher probability of being learned, because the learned item would provide a "blank" period in which additional rehearsal of the previous items could take place. To test the possibility that the learning of a particular item does influence learning of others near it, opportunity for rehearsal was varied. Easily learned items were introduced which, once learned, could serve as "blank" periods during which extra rehearsal could take place.

In addition, rate of presentation was varied. The relatively long exposure time used by Rock (1957) and Estes (1960) allowed much opportunity for rehearsal of each item; a faster rate (more similar to that used in most paired-associate experiments), by cutting down on the time allowed on each item, would be expected to change the opportunity for rehearsal. Actually, performance under a fast presentation rate is more relevant to the issue of all-or-none learning, since with faster rates there is greater control of interitem repetition and rehearsal.

## METHOD

*Subjects.*—The Ss were 24 Yale undergraduates, who participated in the experiment to fulfill a course requirement. Seventy similar pretest Ss were used to standardize the materials.

*Apparatus.*—The material to be learned was typed in capital letters on white adding machine tape, and was presented to S through a 1 × 3 in. aperture in a 12 × 36 in. screen at rates predetermined by E. The S spoke into a small microphone, which he held in his hand during the learning trials. A system of relays, motors, and timers permitted the required variations in presentation time and also the recording of response latencies in units of .01 sec. The E sat behind the screen, hidden from S, recorded S's response on each item, and recorded latencies from two Standard Electric timers.

*Design.*—Each S learned a list of 25 4-letter word pairs. All words in the list were taken from the Thorndike-Lorge lists of the 1,000 most frequently occurring words (Thorndike & Lorge, 1944). Choice of the items was made on the basis of pretesting: lists were learned by a total of 70 pretest Ss, and approximately 50% of the pairs, those most easily learned, were eliminated. The 22 remaining pairs (e.g., WALL-CORN, PAST-FISH) presumably were more homogeneous than the original sample. In addition, three pairs (e.g., FAST-SLOW) were also designed to be easily learned by all Ss.

The order of the pairs in the list was changed from trial to trial as is customary in paired-associate learning. However, to facilitate analysis, the easy items always appeared in Positions 6, 13, and 20, although any particular easy item appeared in different positions on different trials. Each of the other items remained in the same position relative to that of an easy item on all trials. That is, the same three items always appeared just before an easy item (i.e., at Positions 5, 12, or 19), three other items always appeared two items in advance (i.e., at Positions 4, 11, or 18), etc.; again, any one item appeared at all three equally-distant positions (e.g., 5, 12, and 19) on different trials. Three different serial orders of the pairs were possible within these specifications. These orders were alternated, so that on consecutive trials, no pair was (a) in the same position, or (b) adjacent to the same pair as it had been on the previous trial. In addition, there were four other word pairs in the list, two at the beginning, and two at the end. The position of these four pairs varied within these four

end positions, but they never appeared at any other point in the list.

Within the main portion of the list, the items were arranged in four different sequences for different Ss. The position in which each item appeared was chosen randomly in two of these four sequences. The others were simply the first two in reverse order, i.e., the pair which had directly preceded an easy pair now followed an easy pair and was three items removed from it (thus appearing at Positions 9, 16, or 23), etc. This was done in order to balance the pairs appearing at the various positions in the list with respect to item difficulty. The four end items were completely different for the four sequences, and were chosen randomly from the data gathered on the pretest Ss.

The anticipation method was used, and material was presented at two rates. For one group, both members of the word pair were exposed for 1 sec. (fast rate of presentation); for the other group, each pair was exposed for 4 sec. (slow rate). It was felt that these values would ensure the best possible manipulation of rehearsal time on the items. All other values were held constant. The first word of the pair was presented alone to both groups for 3 sec., and there was a 20-sec. intertrial interval.

Twenty-four Ss were assigned randomly to one of the two conditions. Three Ss in each condition learned each of the four item sequences described above.

*Procedure.*—The task was described, and Ss were told that both their responses and the latencies of their responses would be recorded. A ready signal was given 2 sec. before the start of each of the 24 learning trials. After the experiment, Ss were asked to indicate the items which they had learned through simple memorization, and those for which they had utilized a mnemonic aid.

## RESULTS

The 4-sec. group (slow rate) required significantly fewer trials ($M = 8.8$, $SD = 2.3$) to reach a criterion of at least 20 out of 25 pairs correct ($t = 3.93$, $df = 22$, $P < .01$) than did the 1-sec. group ($M = 14.3$, $SD = 4.3$). There were also reliable differences in the mean number of items correct during learning: 4-sec. rate: $M = 463.2$, $SD = 33.8$; 1-sec. rate: $M = 383.2$, $SD = 74.8$ ($t = 3.34$, $df = 22$, $P < .01$).

While the items selected as "easy" were anticipated considerably more often ($M = 266.0$, $SD = 8.9$) than the other items ($M = 194.6$, $SD = 33.3$) for both the 1-sec. and the 4-sec. groups ($F = 42.37$, $MS = 619.5$, $df = 1/38$, $P < .001$), there were no reliable differences in the learning of the other items that could be attributed to proximity to (i.e., either before or after) the easy items or to an interaction of proximity and presentation rate, either in terms of the mean number of errors before the learning criterion (20/25) was reached, or in terms of the mean number of correct anticipations during the first six trials. Thus no evidence was obtained that would seem to preclude treating each item as an independent unit for analysis.

The four additional items placed at the ends of the list were not significantly different in learning rate ($M = 16.2$, $SD = 3.9$) from the regular items ($M = 14.9$, $SD = 4.1$), $F = 1.55$, $MS = 20.3$, $df = 1/44$. Further analyses were done (a) including these four items and (b) excluding them. Since the results were identical in every case, only those analyses in which the end items were included are presented.

The probability of correctly anticipating a response for the first time as a function of trials is shown in Fig. 1 for the first 10 trials for both rate groups. Contrary to one of the major assumptions of an all-or-none position, these curves show a sizeable increase over trials. The obtained linear chi squares (Cochran, 1954) were 21.80 and 41.08 for the 1- and 4-sec. rates, respectively ($df = 1$, $P < .001$ for each group). Only those trials on which the probability is based on more than 100 observations are included in the analysis.

At the slow rate of presentation,

about half of the regular items showed no "breaks," that is, once correct, they were always correct on subsequent trials. However, at the fast rate, there were significantly fewer (94 out of 260) such break-free items ($\chi^2 = 17.50$, $df = 1$, $P < .001$). At both rates there was a greater number of break-free items (1-sec. rate: 26 out of 36; $\chi^2 = 7.12$, $df = 1$, $P < .01$; 4-sec. rate: 28 out of 36; $\chi^2 = 11.11$, $df = 1$, $P < .01$). There were no differences in the number of mnemonics reported by the two groups ($t = .25$, $df = 22$), but in both groups items learned with the aid of a mnemonic showed a significantly greater number correct over the learning trials ($M = 17.2$, $SD = 2.7$) than did the items memorized ($M = 15.1$, $SD = 4.1$), $F = 6.33$, $MS = 57.4$, $df = 1/44$, $P < .025$. In addition, there were fewer memorized items (33 out of 94) among the break-free items ($\chi^2 = 12.7$, $df = 1$, $P < .001$), whereas among the break items, about half were memorized and half learned with the aid of a mnemonic.
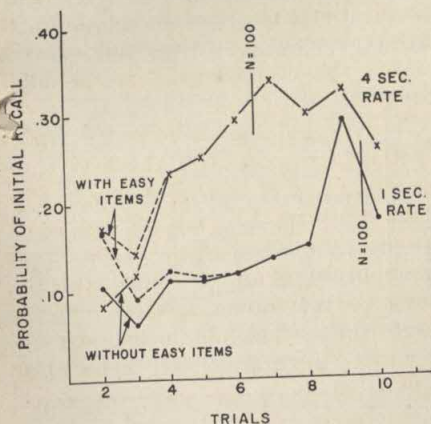


FIG. 1. The probability of correctly anticipating an item for the first time, as a function of trials and presentation rate. (The statistical analysis was done on the portion of the curve that is based on more than 100 observations, as indicated.)
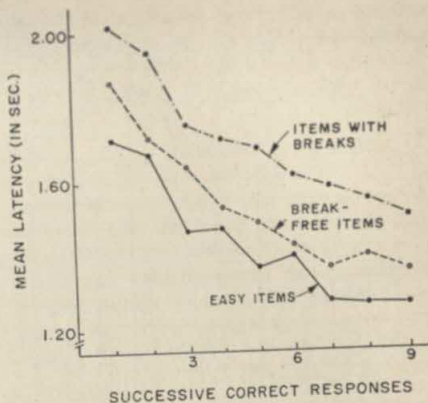


FIG. 2. Mean latency of items on their first correct anticipation, and on successive correct anticipations. (Since there were no differences in latency between the two presentation rates, the data for both groups have been combined.)

There is a similar ordering of items in terms of latency.[3] The easy items exhibit the lowest initial latencies ($M = 1.71$ sec., $SD = .29$); and for those items which, once correct, are correct on all subsequent trials, the initial latencies are lower ($M = 1.87$, $SD = .20$) than for pairs in which there are breaks ($M = 2.02$, $SD = .28$). These differences are significant at the .05 level ($F = 7.85$, $MS = 707.54$, $df = 2/66$). Correlations between the presence of breaks and initial latency were calculated: the point biserial $r$'s were .60 and .58 for the 1-sec. and 4-sec. rates, respectively ($df = 10$, $P < .05$ for both cases). Moreover, there was a tendency, although it did not reach conventional levels of significance, for items which were learned with the aid of a mnemonic to show lower latencies ($M = 1.58$, $SD = .17$) than did those items which

[3] Most of the latency scores fell between 1 and 2 sec. Because of this, and because of the small departure from normality in the distribution of scores, transformation of the data into reciprocals did not seem warranted.

TABLE 1

LATENCY (IN SEC.) OVER BLOCKS OF TRIALS FOR ITEM TYPES AND
RATE OF PRESENTATION

| Items | Rate | Trials 1–3 | | Trials 4–6 | | Trials 7–9 | |
|-------|------|------|------|------|------|------|------|
| | | Mean | SD | Mean | SD | Mean | SD |
| Easy | 1 sec. | 1.57 | .09 | 1.39 | .07 | 1.26 | .01 |
| | 4 sec. | 1.65 | .17 | 1.43 | .06 | 1.27 | .01 |
| Break-free | 1 sec. | 1.73 | .12 | 1.46 | .08 | 1.32 | .02 |
| | 4 sec. | 1.75 | .12 | 1.49 | .02 | 1.42 | .03 |
| Break | 1 sec. | 1.91 | .15 | 1.63 | .04 | 1.52 | .06 |
| | 4 sec. | 1.89 | .13 | 1.71 | .06 | 1.57 | .02 |

were memorized ($M = 1.69$, $SD = .17$), $F = 3.61$, $df = 1/32$, $P < .07$.

In order to assess changes in the strength of association after the point of initial recall, latencies of individual pairs were examined. Figure 2 presents the mean latency of each item on the trial on which it was first correct, and on each successive correct trial. The individual item latencies decrease over trials, and the ordering of the items as it appeared in the initial latencies remains over the course of successive correct trials. The mean latencies for blocks of trials are shown in Table 1, and the analysis of variance is presented in Table 2, indicating that the main effects of blocks of trials and type of item (easy, break-free, break) are significant, that rate of presentation has no significant effect, and that there are no significant interactions.

According to Estes (1960), "to determine whether the behavioral change associated with a decrease in latency is learned on an all-or-none basis, we would need a similar analysis [i.e., his probability analysis on individual items] with some criterion of change in latency as the dependent variable" (p. 221). Following this suggestion, two latency scores, 1.75 sec. and 1.30 sec., were chosen as learning criteria. Figure 3 presents

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSE
LATENCIES

| Source | df | MS | F |
|--------|-----|---------|---------|
| Rate of presentation (A) | 1 | 268.00 | 3.66 |
| Item type (B) | 2 | 3,551.83 | 48.44* |
| Blocks of trials (C) | 2 | 5,907.24 | 80.57* |
| A × B | 2 | 1.62 | <1 |
| A × C | 2 | 8.82 | <1 |
| B × C | 4 | 13.37 | <1 |
| A × B × C | 4 | 33.83 | <1 |
| Error | 36 | 73.32 | |

* $P < .001$.
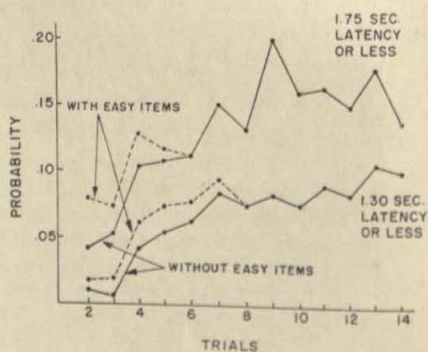


FIG. 3. The probability of achieving two different criterion latencies (1.75 sec. or less, and 1.30 sec. or less) as a function of trials. (All points shown are based on more than 100 observations.)

the probability over trials of achieving these latencies for the first time on each item. The data from Ss run under both rates have been combined, since there were no differences in latency as a function of rate. With these criteria, too, there was a significant increase over trials: the linear chi squares were 60.31, $df = 1$, $P < .001$, and 66.15, $df = 1$, $P < .001$, for the 1.75-sec. and the 1.30-sec. criteria, respectively.

## DISCUSSION

The finding that repetition of unlearned items does in fact increase the probability that they will be "learned" on a subsequent trial is in direct contradiction to the all-or-none assumption that these probabilities are nonincreasing. It is consistent with a strength position, such as that of Hull (1943), indicating that some subthreshold learning is taking place on items before the point of initial recall. The greatest increase occurred beyond the first two trials, which were the ones presented by Estes. In fact, in Estes' (1960) experiment, the probability of initial recall went from .40 on the first trial to .46 on the second; perhaps this trend would have been reliable if the analysis had included further trials.[4]

When the curves for the obviously heterogeneous items (e.g., easy items included) are compared with those for the more homogeneous items, it can be seen that the effect of heterogeneity is to counteract the generally rising trend and to produce an actual decline between Trials 1 and 2. These empirical results support the expectation that the effect of heterogeneity of item difficulty should be to obscure the rise in performance with

[4] This analysis was also done on previously published (Williams, 1961) data from 60 Ss who learned 12-item lists of letter-number pairs by alternate training and testing trials, with 5 sec. exposure per item—conditions similar to those of Rock (1957). These Ss also showed a significant ($P < .01$) increase over trials in the probability of initial recall.

practice, and thus yield an artifact in the direction of results presented to support the all-or-none hypothesis. The greatest effect of heterogeneity in the present experiment is seen in the early trials, and it is possible that the two opposing factors, heterogeneity of item difficulty and repetition, were also operating in the Estes situation. Indeed, it is difficult to conceive of any verbal materials which would be completely homogeneous, especially taking into consideration idiosyncratic sources of difficulty. Even Rock's (1957) experiment, which utilized simple letters and numbers as learning material, showed a selection bias due to differential ease of learning certain of the items (Williams, 1961).

In this experiment it was impossible to identify and exclude items initially guessed correctly instead of recalled. To the extent that S is able to remember which response items he has already used correctly and which he has not, he will be able to increase the probability of a correct guess as the number of unused response terms decreases on later trials. This factor would account for some increase in the probability of initial recall over trials. However, with the relatively long list used in the present experiment, the maximum possible increase in guessing efficiency over the first few trials is not very large. The change in the probability from 25 unlearned items (1/25) to 10 unlearned items (1/10) is only .06. The observed increase in probability during the same trials, however, was .22 for the 1-sec. and .37 for the 4-sec. rate. Moreover, these figures are taken from an examination of the break-free items only, which eliminates most of the items on which the first recall was due to guessing. Furthermore, the actual benefit from guessing was probably considerably less than the allowance made above. The Ss made guesses on only .21 and .18 (in the 1-sec. and 4-sec. groups, respectively) of the failures to respond correctly; the remaining failures were omissions. Also, it is not likely that Ss were able to remember all the response terms

in the list on every trial. For these reasons, it is felt that the effects of increased probability of guessing correctly cannot be used to explain the large increase obtained.

As training proceeds, the number of as-yet-unlearned items decreases, and it might be argued that it is this change in the effective length of the list during learning which accounts for the rise in the probability of correct recall over trials. Further analysis of data from a previously published experiment (Williams, 1961), however, suggests that this is not the case. For 10 $S$s in that experiment, the items responded to correctly in a 12-item list were eliminated on each trial and new ones were substituted for them, so that on every trial $S$ was presented with 12 "unlearned" items. In this manner the effective list length was held constant. These $S$s showed an increase in the probability of initial recall over trials ($P < .001$) just as did the $S$s run in the present experiment.

The finding that latency decreases as a function of practice suggests that after the point of initial recall, and after there is no further change with respect to the recall criterion, a latency measure will reflect a still increasing strength of association. This of course is not directly relevant to the Estes model, which is limited to recall, but it would bear on the formulation of a more comprehensive theory of all-or-none learning.

According to the all-or-none position, all items which are learned are learned to full strength on one trial: "breaks" after initial recall are attributable to forgetting of a learned item or to the fact that the initial recall constituted a "guess" rather than a learned response. The present data indicate that the items which show breaks during learning tend to be the ones which have relatively high latencies on the trials on which they were answered correctly. This suggests that all items which are learned to a criterion of simple recall are not necessarily equal in associative strength—as an all-or-none position implies—either in terms of subsequent recall of the items, or in terms of latency scores.

Recall was greatly influenced by rate of presentation. However, the slower rate, which produced an increase in the probability of recall, did not result in decreased latency. Perhaps a change in rate has a specific effect on latency which counteracts the effects to be expected by strength. For example, $S$ may develop a set or rhythm based on the rate.

Several operations for influencing "strength"—number of trials, rate of presentation, and type of item (easy, break-free, break)—were included in this experiment, and there is also evidence that there is some influence of the availability of a mnemonic aid on strength of association. The utility of an intervening variable such as strength of association depends on demonstrating a relationship between more than one manipulation and more than one measure (Miller, 1959). With the one exception noted above, excellent agreement was found among the several manipulations and measures in the present study, suggesting that response probability and response latency are indeed both measures of the same intervening variable, strength.

## SUMMARY

Subjects learned a paired-associate list by the anticipation method at one of two rates of presentation: each pair was exposed for either 1 sec. or for 4 sec., and the anticipation interval ($S$ term alone) was held constant at 3 sec. The list was composed of 25 simple word pairs. The order in which the pairs were presented varied from trial to trial, but the list was arranged so that each item, though at different positions on consecutive trials, always remained the same distance before (or after) one of three items designed to be easily learned. Recall and latency of response were measured.

1. There were no differences in learning rate among the items that could be attributed to proximity to the easy items, thus indicating that items could be treated as independent units for analysis.

2. The probability of responding correctly to an item for the first time increased as a function of trials. This was true for both a simple recall criterion and for a criterion in terms of latency.

3. Evidence for a selection artifact was demonstrated, in that heterogeneity of difficulty among the items tended to obscure the rise in probability over trials. That is, the easy items were learned on the early trials, and thus the probabilities on those trials were much greater than when only the more homogeneous items were included in the analysis.

4. The latency of individual pairs decreased as a function of successive correct responses.

5. Latency was a function of the type of item: the items chosen to be easily learned exhibited the lowest latencies; and items which, once correct, were correct on all subsequent trials showed lower latencies than did items which contained breaks, i.e., errors after the first correct response.

The results were interpreted as supporting a strength theory.

## References

Brown, J., & Huda, M. Response latencies produced by massed and spaced learning of a paired-associates list. *J. exp. Psychol.*, 1961, **61**, 360–364.

Cochran, W. G. Some methods for strengthening the common $\chi^2$ tests. *Biometrics*, 1954, **10**, 417–451.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–233.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Simley, O. A. The relation of subliminal to supraliminal learning. *Arch. Psychol., N. Y.*, 1933, **22**, No. 146.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Thorndike, E. L. *The fundamentals of learning.* New York: Teachers College, Columbia University, 1932.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Williams, J. P. Supplementary report: A selection artifact in Rock's study of the role of repetition. *J. exp. Psychol.*, 1961, **62**, 627–628.

# INFLUENCE OF A SMALL NUMBER OF PARTIAL REIN-FORCEMENT TRAINING TRIALS ON RESISTANCE TO EXTINCTION [1]

E. J. CAPALDI AND DICK HART

*University of Texas*

The present experiment undertaken within the context provided by the Hull-Sheffield hypothesis was concerned with the effect of a small number of partial reinforcement training trials on resistance to extinction as a function of pattern of reinforcement. According to the generalization decrement hypothesis (Sheffield, 1949) extinction necessarily involves the introduction of stimuli different from those conditioned to the instrumental response in acquisition. The hypothesis also holds that resistance to extinction should decrease to the extent that the response is independent of the control of such newly introduced stimuli.

The application of the general hypothesis to partial reinforcement involves the assumption that reinforcement or nonreinforcement on a particular trial gives rise to distinctive stimuli which become part of the total stimulus complex on the subsequent trial. When stimuli characteristic of nonreinforcement constitute a portion of the stimulus complex on a particular trial and reinforcement occurs, $S$ learns to perform the instrumental response in the presence of cues characteristic of extinction. Consistently reinforced $S$s are denied the opportunity for such conditioning. Evidently, then, at the start of extinction there is less change in the conditioned stimulus pattern for the

partial $S$s; accordingly, the consistent $S$s should extinguish more rapidly.

The generalization decrement hypothesis suggests that resistance to extinction depends upon pattern of reinforcement; our attention is directed to transitions from nonreinforced to reinforced trials (N-R transitions). However, early pattern learning experiments appeared not to support the Hull-Sheffield view. For example, when a moderate number of training trials were employed single alternation of reinforcement (SA), which involves the maximum number of N-R transitions for a given number of training trials, and random training (R), which necessarily involves a smaller number of N-R transitions, yielded about the same degree of resistance to extinction (Capaldi, 1958). Ostensibly even more damaging to the hypothesis, when considerable numbers of training trials were employed SA training was actually followed by lesser resistance than R training (Capaldi, 1958; Tyler, Wortz, & Bitterman, 1953). It remains to be determined whether or not the Hull-Sheffield hypothesis is adequate to deal with patterning effects when only a small number of training trials are employed. Accordingly, the following predictions from the hypothesis were tested in the present study: (*a*) when only a small number of N-R transitions are employed, the typical partial reinforcement effect of increased resistance to extinction will not occur, because the cues characteristic of

nonreinforcement are but weakly conditioned to the instrumental response; (b) when two patterns of partial reinforcement involve the same small number of trials and the same number of reinforcements and nonreinforcements, i.e., identical marginal probabilities, the pattern involving the greater number of N-R transitions will be followed by the greater resistance to extinction.

## EXPERIMENT I

### Method

*Subjects.*—The Ss, approximately 90 days old at the start of the experiment, were 18 male and 21 female experimentally naive albino rats from the colony maintained by the Department of Psychology, the University of Texas. The males and females were distributed equally throughout three groups of 13 Ss each by means of a random procedure.

*Apparatus.*—A straight-alley runway with an overall inside length of $7\frac{1}{2}$ ft. and width of 4 in., enclosed by 8-in. high sides, served as the apparatus. A depression plate at the start end and a photo electric cell at the goal end of the alley served to start and stop, respectively, an electric timer measuring in .01 sec. The distance between the tip of the depression plate and the light beam was 6 ft. 4 in. A sliding inset with two identical-appearing compartments large enough for reward containers, which could be manipulated appropriately on reward and nonreward trials, was situated to the side and at the end of the goal portion of the alley. A guillotine door 1 ft. from the end of the alley could be lowered so as to confine S to the goal box. The entire confinement area was covered with $\frac{1}{2}$-in., hinged hardware cloth. The apparatus was constructed of wood and painted gray throughout.

*Procedure.*—On the initial day Ss were individually housed and deprived of food for 23 hr. On all succeeding days of the experiment Ss were fed for 1 hr. in the home cage. On Days 2 through 7 Ss were handled in groups of 6. On Day 2 food was available during the 1-hr. handling period. On each successive day feeding time outside the home cage was reduced by 15 min. in order to gradually adjust Ss to the 23-hr. deprivation schedule. Immediately following the handling period Ss were given the 1-hr. daily ration in the home cage. On Days 6 and 7 Ss were allowed to explore the runway in groups of 2 each for 30 min., no food being available.

The Ss were fed for 1 hr. in the home cage immediately following exploratory training. Since, in the experimental phase proper, approximately 15 min. were required to administer the daily trials to each S it can be seen that by now each S was being fed at approximately the time its daily trials were due to terminate.

Beginning on Day 8 nine acquisition trials per day for 3 days were given. Group C (consistent) was given food reward in the form of a wet mash following each run, Group SA was rewarded on Trials 1, 3, 5, 7, and 9 of each day. Group R was rewarded on Trials 1, 2, 5, 6, and 7 on Day 1 and on Trials 1, 2, 3, 6, and 7 on Day 2. The pattern on Day 3 was the same as that given on Day 1. Thus, Groups SA and R received equal numbers of reinforcements (R) and nonreinforcements (N). However, while Group SA received four N-R transitions per day, Group R received only one per day. Goal-box confinement was 15 sec. on both N and R trials. The intertrial interval was 15 sec.

Two days (Days 11 and 12) of extinction training were given, 10 trials per day. On the initial day of extinction a single reinforced trial preceded the 10 extinction trials. The length of confinement in the goal box and the intertrial interval remained at 15 sec. If S did not enter the goal box within 75 sec., it was removed from the alley for the 15-sec. intertrial interval. Two such consecutive failures to respond resulted in discontinuance of work with a particular S. An arbitrary time of 75 sec. was assigned for the remainder of the trials. Each S was fed for 1 hr. in the home cage immediately following the termination of its daily trials.

### Results

*Acquisition.*—A repeated measures analysis based on the daily median for each S (data not shown) indicated that the groups did not differ reliably ($F = 1.39$, $df = 2/36$) in acquisition. The interaction between groups and trials was not significant ($F < 1$) nor did performance over trials reach a conventional level of significance ($F = 3.00$, $df = 2/27$, $.05 < P < .10$).

A clear understanding of the data presented in Table 1 requires consideration of the following points. Consider the SA pattern on nine daily trials (R, N, etc.) employed in the present experiment. Trials 2, 4, 6,

and 8 are termed trials following reinforcement (TFR) and Trials 3, 5, 7, and 9, trials following nonreinforcement (TFN). Comparison of the performance of Groups SA and C employing only Trials 2, 4, 6, and 8 is said to involve comparable trials following reinforcement (CFR). A similar comparison employing Trials 3, 5, 7, and 9 is said to involve comparable trials following nonreinforcement (CFN). Of course, the expression CFN is literally incorrect when applied in connection with the C pattern; its general meaning is that the comparison between the partial groups and Group C involves the same ordinal trials in the sequence with the occurrence of nonreinforcement in the partial pattern determining which trials are to be employed for analysis. It should be indicated that in computing TFN for Group R it was deemed advisable to omit those trials which followed a second nonreinforced trial, including only those which followed a single nonreinforcement in the sequence. The tendency was for $Ss$ to run more slowly following two consecutive nonreinforcements.

Table 1 presents performance on CFR and CFN for Groups C and SA and for Groups C and R. The entries for CFR and CFN for Group C, which appear in the table immediately above those for Group SA, were determined employing Group SA as the basis of comparison. Similarly, the CFR and CFN entries for Group C which employed Group R as the basis of comparison appear immediately above those for Group R. It will be noted that for Group C daily differences between CFR and CFN are small regardless of whether Group SA or Group R was employed as the basis for computing these means. In marked contrast it can be seen that Groups SA and R ran more rapidly on CFR

### TABLE 1

MEANS OF LOG RUNNING TIMES ON 3 DAYS OF ACQUISITION FOR GROUPS C AND SA, USING CFR AND CFN, AND FOR GROUPS C AND R, USING CFR AND CFN

| Group | CFR | | | CFN | | |
|---|---|---|---|---|---|---|
| | Day 1 | Day 2 | Day 3 | Day 1 | Day 2 | Day 3 |
| C | 1.12 | .98 | .82 | 1.08 | .94 | .82 |
| SA | .83 | .64 | .56 | 1.06 | .83 | .73 |
| C | 1.08 | .95 | .83 | 1.02 | .94 | .76 |
| R | .90 | .83 | .64 | 1.22 | 1.20 | 1.05 |

than on CFN. Differences between CFR and CFN for Group SA were significant on each of the 3 days of acquisition (Day 1, $F = 24.33$, $P < .01$; Day 2, $F = 7.32$, $P < .05$; Day 3, $F = 5.88$, $P < .05$; $df = 1/12$ in each case). Group R also ran reliably faster on CFR as opposed to CFN on each of the 3 days (Day 1, $F = 8.27$, $P < .05$; Day 2, $F = 15.11$, $P < .01$; Day 3, $F = 20.28$, $P < .01$; $df = 1/12$ in each case).

While Group SA ran more rapidly than Group C on CFN and CFR on each of the 3 days of acquisition, only the differences on CFR reached significance (Day 1, $F = 8.25$, $P < .01$; Day 2, $F = 7.91$, $P < .01$; Day 3, $F = 10.17$, $P < .01$; $df = 1/24$ in each case); the differences on CFN yielded $F < 1$ in each case. These results demonstrate that the faster running of partially reinforced as opposed to consistently reinforced groups is not exclusively a late trial phenomenon (e.g., Goodrich, 1959). Such faster running appears to occur quite early in training; the practice of pooling TFR and TFN has apparently obscured it.

*Extinction.*—As Fig. 1 indicates, the greatest degree of resistance was shown by Group SA, the least by

Group C. A repeated measures analysis over the initial 10 trials, employing the log times on each trial, indicated that the groups differed significantly ($F = 18.06$, $df = 2/36$, $P < .01$). A significant extinction effect occurred over trials ($F = 30.34$, $df = 9/324$, $P < .01$); however, the interaction between Group and Trials was significant at only slightly beyond the 10% level ($F = 1.57$, $df = 18/344$). A factor operating to reduce the significance level of the interaction was the arbitrary ceiling time of 75 sec. Many Ss in Group C and quite a few in Group R failed to respond within 75 sec., while only 1 S in Group SA required the full time. Seven Group C Ss and 3 Group R Ss met the criterion of extinction on the initial day. Differences between the individual means were tested using Duncan's multiple range test (Edwards, 1960). The test indicated that all the groups differed from each other at beyond the .01 level.

A similar repeated measures analysis over Trials 11–20 of extinction indicated that the Group ($F = 13.52$, $df = 2/36$, $P < .01$) and Trials ($F = 60.65$, $df = 9/324$, $P < .01$) differences were significant as was the interaction ($F = 3.66$, $df = 18/324$, $P < .01$). The significant interaction term is not too meaningful over this block of trials since it merely serves to indicate that the SA group was approaching the 75-sec. limit, something Groups C and R had largely achieved earlier. Almost all Ss had reached the criterion of extinction at the end of the second day. Duncan's test indicated that Groups SA and C differed beyond the .01 level while the C vs. R and the SA vs. R comparisons were significant beyond the .05 level.

## EXPERIMENT II

The results of the initial experiment are consistent with deductions from the aftereffects hypothesis. However, as an examination of the predictions from the hypothesis which were considered earlier will make clear, the aftereffects view implies that following SA training the usual partial reinforcement effect can be obtained while at the same time Group R and Group C are about equally resistant to extinction. It was this spectrum of results that Exp. I was designed to obtain. Apparently a wrong "guess" was entertained and slightly too many N-R transitions were given in the case of Group R, thus allowing this group to show the usual partial reinforcement effect. In order to obtain the desired results, the number of training trials was reduced in Exp. II from 27 to 18. The experimental procedure was the same as that employed in Exp. I except for the changes noted below.

### Method

*Subjects.*—The 39 naive albino rats were 110 days old at the start of the experiment. The 15 males and 24 females were distributed equally throughout the three groups by means of a random procedure.

*Procedure.*—Since only 2 days of acquisition training were to be given, instead of 3, 1 additional day on the feeding schedule was employed. The patterns of reinforcement employed on Days 1 and 2 of acquisition in the preceeding experiment were employed
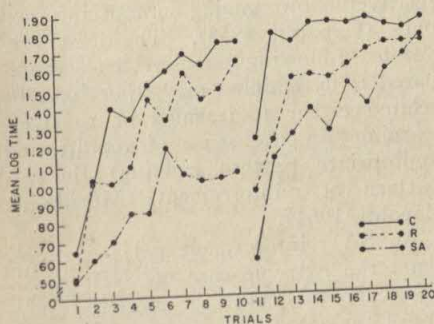


FIG. 1. Mean log running time for each group on each of the 20 trials of extinction in Exp. I.

here on those days. On the day following the final acquisition trial, all *S*s were given a single reinforcement followed by 15 extinction trials, with no further training being given.

## Results

*Acquisition.*—The acquisition results were highly similar to those reported in the initial experiment. Accordingly, these data are not presented here and the results of a particular statistical analysis will be mentioned only if it possesses special relevance or if the significance level deviates from that reported earlier.

The repeated measures analysis based on all trials yielded an $F$ of 3.03 for Groups $(.05 < P < .10, df = 2/36)$, of 1.20 for Trials (not significant for 1/36 $df$), and of less than 1.00 for the interaction between Group and Trials. Unlike the results of Exp. I, differences between Groups SA and C on CFN on Day 2 were significant $(F = 4.95, df = 1/24, P < .05)$, Group SA running more rapidly. From the two analyses reported, it can be seen that the differences between Groups SA and C were greater in the present experiment than in the initial one. Differences between Group C and Group R were of about the same order as previously reported.

*Extinction.*—As Fig. 2 shows, Groups R and C did not differ appreciably in extinction while Group SA appears to show the usual partial reinforcement effect. The log times on each of the 15 trials were summed and an analysis was performed which yielded an $F$ of 9.68, which for 2/36 $df$ is significant beyond the .01 level. Duncan's test indicated that in order for Groups R and C to differ at the .05 level the shortest significant range required was a value of 2.743. The obtained value of 1.180 fell far short of significance. Duncan's test further indicated that differences between
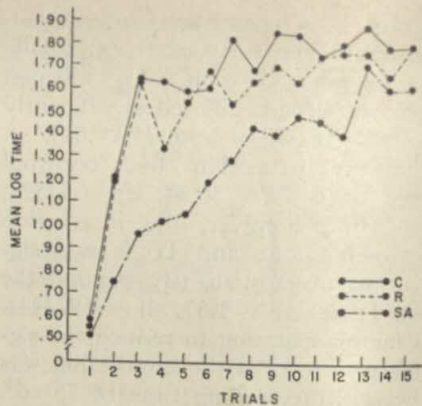


FIG. 2.  Mean log running time for each group on each of the 15 trials of extinction in Exp. II.

Group SA and Groups R and C were significant well beyond the .01 level.

## DISCUSSION

The major finding of the present experiments seems to be that when relatively few training trials are employed SA patterns result in greater resistance to extinction than R patterns. This result can be obtained when the R group is more resistant and when it is about equally resistant to extinction as compared with Group C. It should be noted that current evidence indicates that the tendency of SA reinforcement to result in equal or greater resistance than R patterns occurs prior to the appearance of pattern running, i.e., relatively rapid running on reinforced trials, relatively slow running on nonreinforced ones (Capaldi, 1958). The lesser resistance following extensive SA as compared to R training may, therefore, be related either to learning or to over-learning (in the sense of manifesting appropriate pattern running) the SA pattern of reinforcement (Murillo & Capaldi, 1961).

An SA reinforcement pattern represents the extreme case of N-R transitions. In this sense the present results are of the same genre as those reported by Grosslight, Hall, and Murnin (1953). In that experiment human *S*s given N-R

patterns were found to be more resistant than those given R-N patterns who were, in turn, more resistant than those given R-R patterns. These results are similar to those of Exp. I of this report. The finding reported in Exp. II, that following a small number of trials C and R patterns result in about the same degree of resistance, confirms the results of an earlier investigation by Amsel (1958) who also employed small numbers of trials in connection with C and R patterns.

The present results can be understood in terms of the Hull-Sheffield hypothesis. As previously indicated, variations involving SA and R patterns in connection with moderate to extensive numbers of training trials appear to fail to support the generalization decrement view. As is usual in cases of this kind, movement in one of two general directions is permissible. On the one hand, the hypothesis in question may be abandoned on the ground that attempts to modify it even if successful may not prove to be especially profitable. On the other hand, attempts to modify the hypothesis by considering alternatives which involve logical extensions of it may ultimately prove to be a fruitful course of action. It is our impression that the latter alternative is deserving of serious consideration primarily because pattern learning experiments involving transfer so clearly indicate that the reinforcement outcome of the previous trial results in a modification of the stimulus complex on the subsequent trial (Bloom & Capaldi, 1961; Capaldi & Senko, 1962). Accordingly, it seems reasonable to assume in view of the predictive success of the hypothesis in connection with small numbers of trials that the current inadequacies of the Hull-Sheffield view invite nothing so much as more diligent theoretical analysis and further experimental activity.

## SUMMARY

Two experiments were performed in which rats were trained to traverse a straight alley under either continuous, irregular, or single alternation of reward. In the initial experiment, which employed 27 training trials, the continuous group was found to be least resistant to extinction. The single alternation group was found to be more resistant than the irregular one. Previous experiments have shown that following a moderate number of training trials the alternation and irregular groups are about equally resistant to extinction, while following considerable training the alternation group is less resistant than the irregular one. In the second experiment, extinction training was given following only 18 training trials. While the irregular and continuous groups failed to differ in extinction, the single alternation group showed the typical partial reinforcement effect. The findings were discussed in connection with the Hull-Sheffield aftereffects hypothesis.

## REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BLOOM, J., & CAPALDI, E. J. The behavior of rats in relation to complex patterns of partial reinforcement. *J. comp. physiol. Psychol.*, 1961, **54**, 261–265.

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51**, 367–371.

CAPALDI, E. J., & SENKO, M. G. Acquisition and transfer in partial reinforcement. *J. exp. Psychol.*, 1962, **63**, 155–159.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–64.

GROSSLIGHT, J. W., HALL, J. F., & MURNIN, J. Patterning effect in partial reinforcement. *J. exp. Psychol.*, 1953, **46**, 103–106.

MURILLO, N. R., & CAPALDI, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, **61**, 345–349.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *Amer. J. Psychol.*, 1953, **66**, 57–65.

# CONTIGUOUS CONDITIONING [1]

## W. J. BROGDEN

### University of Wisconsin

The status of reinforcement as a necessary condition for learning has been the focus of much of the controversy between proponents of different theories of learning (Kimble, 1961). It is difficult to take the opposing view when the proponent of reinforcement theory falls back on secondary reinforcement. This latter phenomenon is of course an example of acquisition, of the very thing for the occurrence of which reinforcement is proposed as an essential condition. Nevertheless, it has been argued that the evidence contrary to reinforcement as a necessary condition provided by experiments of latent learning (Tolman & Honzik, 1930) and sensory preconditioning (Brogden, 1939) is not valid because of the possible involvement of primary or secondary reinforcement factors in these phenomena. Contiguous conditioning appears to be completely independent of variables or conditions currently related either to primary or secondary reinforcement. In the experiment to be reported, contiguous conditioning is represented by cage-turning responses of cats in the rotator to a tone CS. This CR is dependent upon a prior conditioning procedure during which each occurrence of the cage-turning response resulted in the sounding of the tone. Initial tests of the tone prior to conditioning elicited no cage-turning responses. Thus contiguous conditioning appears to

be dependent solely upon the temporal contiguity of stimulus and response.

## PROCEDURE

The design of the study to be reported includes one experimental treatment and one control treatment, with the possibility of a second control treatment. It is based upon the positive results of five preliminary studies of contiguous conditioning in which experimental conditions were varied unsystematically and in which there were no control procedures.[2]

The 23 Ss were kittens, approximately 60 days of age at the start of the experiment. They were obtained as six litters, with each litter split between Group E (experimental) and Group C (control) but with random assignment of individual Ss within litters to the two conditions so that N for Group E was 11 and for Group C was 12.

The rotator (Brogden & Culler, 1936) provided cage-turning as the response-to-be-conditioned. The CS was a 1000-cycle pure tone whose intensity was 60 db. above .0002 dyne/cm². The duration of the tone was 4 sec. for all conditions and the duration of each test period was 10 min.

All Ss were given five procedures in succession: (a) Two test periods of adaptation to the rotator. (b) One test period during which two trials of the tone were given when S was quiet to test neutrality of the tone in evoking cage-turning responses; none of the Ss made responses to these test trials. (c) Successive test periods until a final test period was reached that provided a cumulative total of 30 or more cage-turning responses; Group E involved presentation of the tone CS with each occurrence of the cage-turning response and there was no presentation of the tone to Group C. (d) Twenty test trials for contiguous conditioning, each consisting of presentation of the tone alone when S had been quiet for 30 sec. or more. (e) Instrumental shock-avoidance training to the tone CS at the rate of 20 trials per test period until a criterion of 18 CRs was attained. The dura-

[2] Richard F. Thompson collaborated with the author in the preliminary studies.

tion of every response during the training and test procedures was recorded for each $S$ to the nearest .1 sec. by an electronic timer operated by $E$. Groups E and C were treated identically except for the conditioning or training procedure, during which Group E received the tone CS upon making each of the 30 or more cage-turning responses whereas Group C received no tone.

## RESULTS

The second control procedure would have involved a variation of the training procedure to include 30 presentations of the tone when $S$ was quiet in addition to the occurrence of 30 or more cage-turning responses independent of the tone. It was decided in advance that this control would be completed only if the experimental $S$s showed greater activity or responsiveness than the control $S$s, thus indicating a differential effect of the tone during the training procedure. Therefore initial analyses were made of the number of test periods of training procedure including the criterial period, total responses, and frequency of responses per test period. The means are, respectively, 3.36 ($\sigma_m = .54$), 35.80 ($\sigma_m = 1.71$), and 15.90 ($\sigma_m = 3.19$) for Group E and 4.67 ($\sigma_m = 1.36$), 37.82 ($\sigma_m = 1.87$), and 15.91 ($\sigma_m = 3.31$) for Group C. Since not one of the $t$ values for the three differences exceeds 1, there is no evidence of significantly greater activity on the part of Group E. The second control procedure, therefore, was not conducted.

Comparison was also made of the duration of response during the training procedure. The mean for Group E was 6.74 sec. ($\sigma_m = .88$ sec.) and for Group C was 4.68 sec. ($\sigma_m = .52$ sec.). The difference of 2.06 sec. in favor of a greater duration of response by Group E has a $t$ value of 2.02 that is significant at the 10% level but not at the 5% level. Because the duration of response appears to increase

as a function of trials, mean duration was computed for each successive block of 10 trials for the first 30 trials for each $S$. These means are 4.89, 5.87, and 8.49 sec. for Group E and 4.07, 4.61, and 4.97 for Group C. In separate analyses of variance of repeated measures, Trial Blocks has an $F$ value significant at the 5% level for Group E but not for Group C. An overall analysis of variance shows Trial Blocks to be a significant source of variation, but Groups and the interaction of Groups and Trial Blocks are not significant.

Because of the progressive increase in duration of response over trials, and the possibly longer response duration of the experimental $S$s during the training procedure, analyses of the responses during the test procedure were made. These responses are the ones made in between the 20 test trials of the tone CS. The mean frequency of response per test period is 14.99 for Group E and 14.94 for Group C. The difference is not significant nor are these means significantly different from the comparable means of both groups for the training procedure. The mean duration of response is 5.78 sec. for Group E and 5.21 sec. for Group C. The difference is not significant nor are these means significantly different from those of either group during the training procedure. There are no trend effects in response duration for either group during the test procedure.

Of the two measures available to provide a test of contiguous conditioning, the shock-avoidance training procedure measures showed no significant difference between groups. The test trials of the tone CS, presented when $S$ had been quiet for 30 sec. or more, do provide evidence of contiguous conditioning. These data are presented in Table 1. The dif-

| Number of Responses | Number of Animals | |
|---|---|---|
| | Experimental | Control |
| 0 | 2 | 10 |
| 1 | 3 | 0 |
| 2 | 2 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 1 |
| 6 | 1 | 0 |
| 9 | 1 | 0 |
| 12 | 1 | 0 |
| Mean | $3.46(\sigma_m = 1.19)$ | $0.58(\sigma_m = 0.39)$ |

ference in frequency of CR of 2.01 in favor of Group E has a $t$ value of 2.13 and is significant at better than the 5% level. Analysis of variance of these data shows a significant $F$ (7.44) for Group E vs. Group C, but no significant $F$ values for Litter or the interaction of Treatment and Litter. Mean duration of CR is 3.94 sec. for Group E and is 7.24 sec. for Group C. The difference is not statistically significant.

## DISCUSSION

The significantly greater frequency of CR to the test trials of the tone CS by the experimental $S$s over the control $S$s is evidence of contiguous conditioning. These results coupled with the evidence of similarity in frequency and duration of extraneous responses during the testing procedure, establish the validity of contiguous conditioning. It can be argued that the increase in duration of response is also evidence of contiguous conditioning. Such an argument requires the following assumptions: (a) that the cage-turning response elicited by unknown stimuli is of constant duration over training trials as is the case for Group C; (b) that a cage-turning CR is formed during the course of contiguous conditioning training; (c) that the CR increases progressively in magnitude as training increases; and (d) that the mag-

nitude of the cage-turning CR is added to the magnitude of the cage-turning response elicited by unknown stimuli. This hypothesis is consonant with the experimental evidence. Other hypotheses are not. Facilitating action of the tone to increased duration of the response progressively is unlikely. Facilitation does not occur as a trend over trials nor under the time relations of the present study. The tone occurs after the response has started, the delay being equal to the reaction time of $E$. There is little support for the hypothesis that the experimental $S$s learned to "turn on the tone," since there are no group differences during the training phase for total responses, number of test periods, or frequency of response per test period.

The progressive increase in duration of the cage-turning response by the experimental $S$s during the training phase may be evidence of contiguous conditioning in addition to that provided by the test phase. In any case, it does not appear to interfere with interpretation of the test phase data. Thus, the conclusion stands that contiguous conditioning is a function of the contiguity of stimulus (tone) and response (cage-turning) during the training procedure. It follows that contiguity is a sufficient condition of learning. Whether it is a necessary condition is another matter. In considering contiguity versus reinforcement as necessary conditions of learning, it should be noted that any experimental test of reinforcement would appear to confound the reinforcement operation with contiguity of stimulus and response. Such confounding does not occur in the test of contiguity provided by the present experiment.

In further consideration of contiguous conditioning, it should be noted that the CR it produces is weak relative to the CR of standard conditioning procedures involving reinforcement. An important factor may be the backward time relations inherent in contiguous conditioning training. The CS cannot be presented until after the response has started. Even though there may be discovery of training conditions for contiguous condi-

tioning that provide a greater strength of CR, it is unlikely that the CR will be anywhere near as strong and stable as the CR produced by conditioning procedures in which contiguity and reinforcement are confounded. If, however, efficiency is considered in terms of number of training trials required to produce a CR, then the efficiency of contiguous conditioning appears to be high. Contiguous conditioning was obtained with as few as 10 training trials in the preliminary studies. Evidence that sensory preconditioning occurs with 1 or 2 trials of preconditioning and attains a maximum with 4 trials (Hoffeld, Kendall, Thompson, & Brogden, 1960) suggests a similar relation between amount of training and magnitude of contiguous conditioning. When contiguity of stimulus and response are confounded with reinforcement operations in the more standard conditioning procedures, contiguity may produce learning in the early trials, with reinforcement functioning in later trials to fixate and strengthen the connection established by contiguity.

Contiguous conditioning and sensory preconditioning are similar phenomena in that learning occurs apparently without reinforcement. In contiguous conditioning, the training procedure involves a response for which the stimulus is unknown and a stimulus for which the response is not known. In sensory preconditioning, the training procedure involves two stimuli to each of which the response is unknown. Since we must assume the antecedent occurrence of a response following the presentation of a stimulus, both the procedures for contiguous conditioning and sensory preconditioning have serious lacks in precision of experimental manipulation and control. Speculation beyond an assumed stimulus-response relationship established during training and based upon contiguity, that persists in the organism to the testing procedure, appears fruitless until there are experimental operations to identify and manipulate the presently unknown variables of the training procedures.

Perhaps we are asking the wrong questions when we concern ourselves with what condition or conditions are necessary and sufficient for learning to occur. Another approach to our understanding of learning follows from the assumptions that learning occurs almost continuously in organisms of many species, and that in experiments on learning, Ss will always learn, but not necessarily in terms of the behavior for which the experiment was designed. Then the major problem of learning becomes the identification and attainment of experimental control of variables that produce different kinds of learned behavior and of variables that reduce the probability of learning other than that for which the experiment is designed. From this point of view contiguous conditioning may interfere seriously with investigation of the more stable varieties of learning. If contiguous conditioning does occur with only a few trials in which a given stimulus and response are contiguous, the opportunity for establishing a number of contiguous CRs will occur in virtually every experiment on learning. Much of the variance in learning experiments may be due to the occurrence of contiguous CRs. If this is so, then the primary contribution of contiguous conditioning to the general study of learning will come from discovery of conditions that reduce its occurrence, rather than in demonstrating that contiguity alone is a sufficient condition for learning.

## SUMMARY

Contiguous conditioning is represented by cage-turning CRs of cats to a tone CS. This CR is dependent upon a prior conditioning procedure during which each occurrence of the cage-turning response results in the sounding of the tone. Initial tests of the tone prior to the conditioning training elicited no cage-turning responses. A control group, not given the tone CS, made the same number of responses in the rotator prior to the test for contiguous conditioning that the experimental group made during its training procedure. Tests with the tone CS presented when the S had been quiet for 30 sec. or more were given to all Ss. The frequency of CR of the experimental group was significantly greater

than the frequency of response of the control group.

The evidence of contiguous conditioning demonstrates that contiguity of stimulus and response is a sufficient condition for learning. The results are discussed relative to reinforcement as a necessary condition of learning, to the inherent confounding of any reinforcement operation with contiguity of stimulus and response, to sensory preconditioning, and to general theoretical considerations of learning.

### ' REFERENCES

BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.

BROGDEN, W. J., & CULLER, E. Device for the motor conditioning of small animals. *Science*, 1936, **83**, 269–270.

HOFFELD, D. R., KENDALL, S. B., THOMPSON, R. F., & BROGDEN, W. J. Effect of amount of preconditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1960, **59**, 198–204.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

TOLMAN, E. C., & HONZIK, C. H. Introduction and removal of reward, and maze performance in rats. *U. Calif. Publ. Psychol.*, 1930, **4**, 257–275.

# LEARNING OF SIMPLE STRUCTURES [1]

## GEORGE MANDLER AND PHILIP A. COWAN

*University of Toronto*

In the exploration of variables that influence human learning, more attention has recently been paid to the experimental correlates of such concepts as schemas, strategies, organizations, and plans. However, little systematic work is available on the effect of such structural variables on relatively simple human learning. One possible approach was recently presented by DeSoto (1960) in a study ostensibly designed to explore social psychological variables.

DeSoto showed that *S*s learned a social structure, i.e., correctly identified such relations as "*x* influences *y*," in fewer trials when the relations followed the usually encountered properties (such as symmetry, transitivity, and completeness) of these relationships. If these "social relations" are learned according to logical and quasilogical schemas, similar effects should be demonstrable when the relations to be learned, and the members of the set among which the relations operate, are neutral, i.e., not influenced by prior social expectations.

Given a set of three elements (A, B, and C) and an attribute R which may or may not be present with any pair of elements, there are 16 possible paradigms (or structures) ranging from R occurring with no pair to R occurring with all of the six possible ordered pairs. If *S*s are given the task to learn or identify those pairs

which do and do not have the R attribute, two variables might influence speed of acquisition: the number of R (and non-R) attributes in a set of six pairs; and the logical and quasilogical relations within these sets. On the basis of number of attributes alone, sets with one R and five non-R attributes should be easier to discriminate than sets with two and four or three and three R and non-R attributes. However, comparisons of different sets with equal number of attributes should reveal the operation of structural factors. The structure of a set describes the relations among the three elements. These relations are defined by specifying which pairs of elements do or do not have the attribute, thus involving both ordering and attribution. There are 16 possible structures under these conditions; they are shown in Fig. 1. We have arbitrarily labeled the three elements A, B, and C, and indicated the presence of the R attribute by a directed arrow; the direction of the arrow indicates the ordering of the pair.[2] Non-R attributes are inferred from the absence of such an arrow, and in written description by the absence of R. Thus for example, Structure 2 may be described as ARB, AC, BA, BC, CA, CB; Structure 7 as ARB, BRA, BRC, AC, CA, CB. We would expect, for example,

[2] In terms of graph theory (Harary & Norman, 1953) this structure may be characterized as a directed graph of Type 2 with three points and with the two relations denying each other. A directed graph with three points has six possible lines, and in the case of the R and non-R relations a definition of these six lines defines the graph.
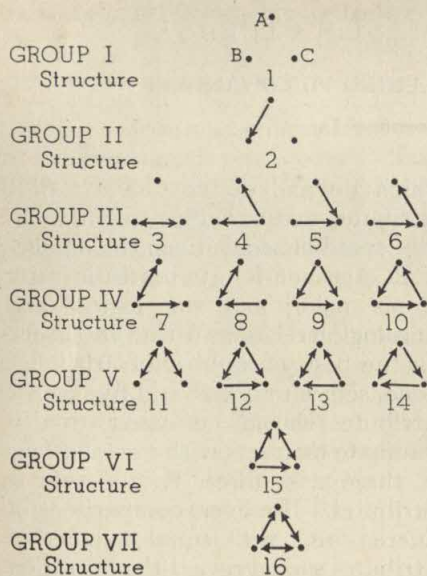
FIG. 1. Diagrammatic presentation of the 16 structures. (Structures within groups have the same number of attributes—checkmarks. In the diagrams an arrow indicates the presence of a checkmark for the two elements ordered by the arrow.)

that a structure where both the AB and BA pairs are associated with the attribute would be learned faster than one where such symmetry is not present.

## METHOD

*Task.*—In the present study we investigated the speed with which Ss learn the presence or absence of an attribute (R and non-R) in structures consisting of three elements. Three CVC syllables were used as the three elements and the presence or absence of a checkmark to indicate the R and non-R attribute. In a trial, Ss were presented with six cards. A particular card had on the face side one of the six pairs of syllables separated by a space, and on the reverse side either a checkmark (attribute) or not. The Ss task was to learn which cards had checkmarks and which did not.

Each S was given a deck of 30 cards (representing five trials of six pairs on a particular structure) and three answer sheets with 30 spaces on each numbered 1 to 30, 31 to 60, 61 to 90, respectively, i.e., Trials 1 to 5, 6 to 10, 11 to 15. The order of syllable pairs with-

in any one trial of 6 cards was randomized with the restriction that in adjacent trials no cards with the same pairs would follow each other. The Ss were required to look at the face of each card, indicate on the answer sheet whether they thought that the reverse side had a checkmark on it by either placing a checkmark or a straight line next to the appropriate number, then turn the card over, look at the reverse, and then proceed to the next card. They were paced at a speed of 7 sec. per card.

Two sets of three CVC syllables were used with half of the Ss assigned to each set. The CVCs were low-association value syllables with the additional restriction that no letter of the alphabet was used more than once in a set of three syllables. Since no significant differences were associated with the two sets of syllables, no further reference will be made to this aspect of our design. The two sets of syllables were: XUR, ZIC, GYQ; and KEF, ZUV, QIJ.

*Subjects and instructions.*—The Ss were 192 female students in the introductory general psychology course at the University of Toronto. Twelve Ss were used for each of the 16 structures shown in Fig. 1. The Ss were seated in a large auditorium and the 192 packs and answer sheets were randomly distributed to Ss.

The Ss were told what a CVC nonsense syllable is and were shown the two sets of three syllables. They were told that every card would have on its face two of the three syllables in various combinations. The operative part of the instructions read:

> On the front of some cards there may be a checkmark like this (demonstrated on blackboard); on the back of others there may not be a checkmark. Your pack may have a checkmark on every card, or it may have no checkmarks on any card, or your pack may contain some cards with checkmarks on the back, and other cards without them. . . . You are to find out, or learn, which pair or pairs of nonsense syllables have checkmarks and which pair or pairs of nonsense syllables do not have checkmarks. . . . The cards that have the same pair of syllables on the front will appear over again throughout the pack. . . . You are to learn which pairs of nonsense syllables are associated with a checkmark, and which pairs are not.

The Ss were paced by E telling them to look at the front of a card, mark the answer, look at the back, look at the next card, and so forth.

*Analysis.*—Sixteen different structures presented as six pairs made up from three nonsense syllables were learned by 12 *S*s for each structure. The dependent variable was the number of correct anticipations of the presence or absence of checkmarks. All *S*s were given 15 trials with the order of cards repeated after 5 and 10 trials (30 and 60 cards), respectively. In a subsequent analysis the number of checkmarks per trial was determined regardless of the correctness of the response.

## RESULTS

In discussing the results of the acquisition data, two considerations should be borne in mind. First, there are two independent variables: the number of R and non-R attributes in a structure, and the comparison among structures with the same total number of attributes (or checkmarks). Second, certain structures are symmetrical with respect to the R non-R attributes. Thus, Structures 1 and 16 are identical except for the replacement of R with non-R; the same holds for Structures 2 and 15, 3 and 11, 4 and 12, 5 and 13, and 6 and 14.

*Number of attributes.*—The prediction was that structures will be learned in the following order: Groups I and VII, Groups II and VI, Groups III and V, and Group IV in descending order of discriminability of R and non-R attributes. The mean numbers of correct responses (out of 90) for these four groupings were 87.83, 77.17, 67.12, 61.12 in the predicted order. All differences are significant at the .01 level or better.

*Structural differences.*—The major question to be asked about the structural variables is whether there are significant differences in the acquisition of different structures when the total number of R and non-R attributes is held constant. This problem is examined by looking at differences in acquisition within the groups which contain more than one structure (Groups III, IV, and V). It will be

recalled that Groups III and V are symmetric with reference to R and non-R.

Figure 2 shows the acquisition curves for all structures. Tables 1 and 2 show the mean number of correct responses per trial for all 16 structures and the relevant analysis of variance. The analysis shows highly significant effects for Structures and Trials, and no significant interaction between these two effects. To examine the ordering of structures within groups Tukey's gap test was used to segregate scores at the .05 level of significance.

For Group III the gap test on the scores of Table 1 segregates the structures into the following clusters, with $>$ indicating a .05 gap: $3>4$, $5>6$. For Group V the ordering is: $11>12>13, 14$. In Group IV the following clusters were obtained: $8>9$, $7>10$. Thus, significant differences obtain among structures within groups.

In comparing pairs of structures symmetric with respect to checkmarks and noncheckmarks we note that for five such pairs of structures (Pairs 2 and 15, 3 and 11, 4 and 12, 5 and 13, 6 and 14) the structure with the fewer checkmarks is learned faster and the difference between these pairs is significant at the .05 level for all except Pair 6 and 14.

*Probability matching.*—The results presented have only considered performance in terms of number of correct responses, with correctness defined as a response appropriate to the information on the face of the cards, i.e., the degree to which *S*s correctly learned the event information provided for them. In contrast to this analysis of *event matching*, the data can also be analyzed in terms of *probability matching*, i.e., the degree to which *S*s' behavior conforms to the frequency or percentage of re-
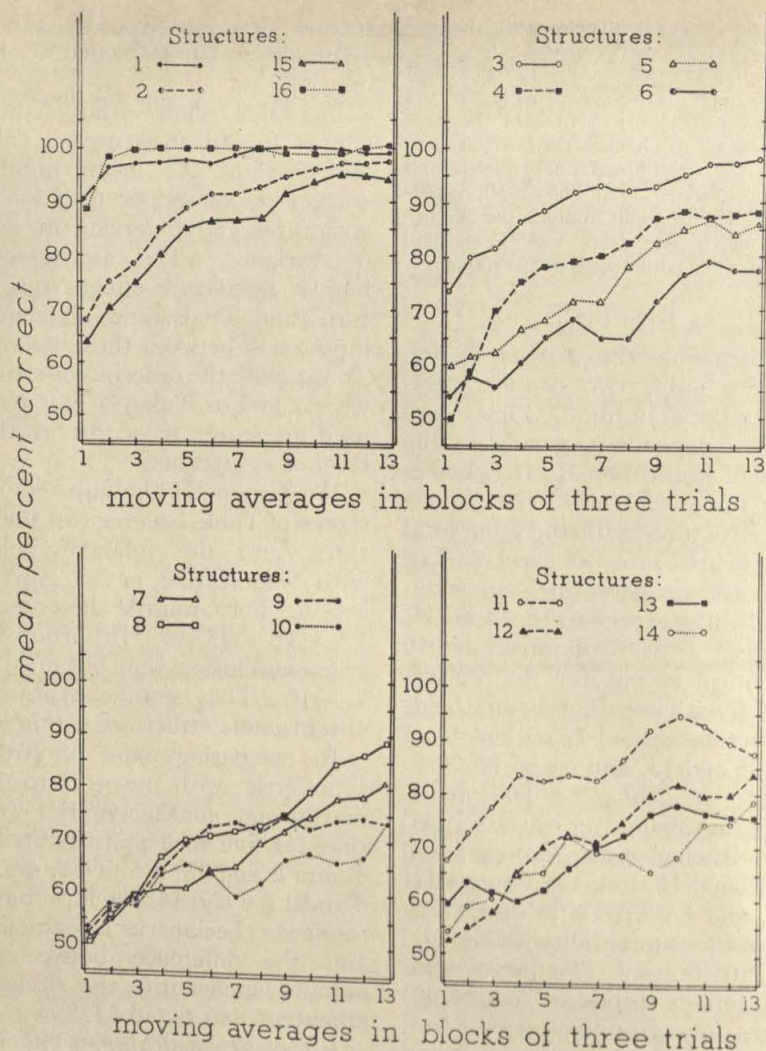
FIG. 2. Acquisition curves for all 16 structures, showing percentage correct responses as a function of blocks of three trials in moving averages.

sponses (checkmark or noncheckmark) required in the task without any attention to the "correctness" of their responses. The data are combined for each of the seven groups which vary from zero to six in the *number* of checkmarks required in each trial of six cards. Figure 3 shows frequency of checkmark or noncheckmark responses for all seven groups. It is obvious that even from the first block of three trials Ss' behavior follows the probability structure, and by Block 5 (Trials 5–7), mean performance is at or near the asymptote expected from sheer probability matching behavior.

When probability matching (frequency only) and event matching (correct responses) are compared for

## TABLE 1

### MEAN NUMBER CORRECT PER TRIAL OF SIX RESPONSES

| Structure | Mean Number Correct |
|-----------|---------------------|
| 16 | 5.86 |
| 1 | 5.85 |
| 3 | 5.37 |
| 2 | 5.26 |
| 15 | 5.03 |
| 11 | 5.00 |
| 4 | 4.54 |
| 5 | 4.46 |
| 8 | 4.30 |
| 12 | 4.27 |
| 13 | 4.13 |
| 9 | 4.04 |
| 6 | 4.03 |
| 14 | 4.02 |
| 7 | 4.02 |
| 10 | 3.82 |

Note.—Lines between structures indicate significant differences at the .05 level by Tukey's test for gaps and stragglers.



GROUP: I ○——○ (X̄ not checked)
II ○——○ (X̄ not checked)
III ○——○ (X̄ not checked)
IV ○——○ (X̄ not checked)
V •----• (X̄ checked )
VI ▲--▲ (X̄ checked )
VII ▼--▼ (X̄ checked )

FIG. 3. Mean number of responses per trial as a function of moving blocks of three trials. (The more frequent response is reported, i.e., checkmarks for Groups V, VI, and VII and noncheckmarks for Groups I, II, III, and IV. Responses were scored regardless of whether they were correct or not.)

Groups II to VI (for Groups I and VII, probability and event matching curves are, of course, identical) it is evident that even when event matching proceeds slowly, probability matching reaches asymptotic or near asymptotic behavior early in the task. If the groups symmetrical with respect to R and non-R relations are examined, it appears that probability matching is even more pronounced in

those cases where event matching proceeds more slowly. Thus, event matching in Structure 2 is faster than in Structure 15, yet probability matching is at 90% of asymptote by Block 2 in Structure 15 (Group VI), while it does not reach that level until Block 4 for Structure 2 (Group II). Similarly, Group V reaches 90% of asymptote by Block 2, while Group III does not reach that level until Block 7. In all groups other than Group III, however, the level of probability matching reaches 90% of asymptote by Block 4, i.e., after 6 trials or 36 presentations of individual cards.

## TABLE 2

### ANALYSIS OF VARIANCE OF NUMBER CORRECT PER TRIAL

| Source of Variance | df | MS | F |
|--------------------|------|-------|--------|
| Total | 2879 | | |
| Between Ss | 191 | 13.18 | |
| Structures (S) | 15 | 82.62 | 11.36* |
| Error | 176 | 7.27 | |
| Within Ss | 2688 | 1.30 | |
| Trials (T) | 14 | 69.95 | 75.22* |
| T × S | 210 | 1.13 | 1.22 |
| Error | 2464 | .93 | |

* P < .001.

## DISCUSSION

The between-groups comparisons—differentiated by absolute number of R and non-R attributes—are clear-cut and predictable from a mere consideration

of discriminability. Obviously, a task which requires $S$ to assign one checkmark vs. five noncheckmarks to six stimuli (pairs) is easier than one which requires discrimination of two and four, or three and three checkmarks and noncheckmarks. The number of stimuli to be discriminated increases and so does the difficulty of the task.

In accounting for the effects within groups, i.e., differences among structures of equal discriminability, we distinguished among four relations that may be differentiated within any one structure. These were *symmetry* (if ARB then also BRA), *transitivity* (if ARB and BRC, then also ARC), *common origin* (if ARB then also ARC), and *common goal* (if ARB then also CRB). The assumption is made that it is this kind of plausible logical and quasilogical reasoning (or use of transformation rules) that facilitates learning of these structures. In Table 3 are presented the number of these relations found within Structures 3 to 14, i.e., all those where there is more than one structure within a group. All relations, whether defined by the presence or the absence of the checkmark attribute, are included. For example, Structure 3 (cf. Fig. 1) has three symmetry relations: AB and BA (no checkmark), BRC and CRB (checkmark), AC and CA (no checkmark).

The order in which structures within groups are learned as a function of number and type of relations may now be considered. For Groups III and V the order of acquisition of structures complementary with respect to R and non-

R is identical, i.e., 3, 4, 5, and 6 in III; and 11, 12, 13, and 14 in V. From Table 3 it appears that symmetry is, and could reasonably be expected to be, the most powerful single relation. Structures 3 and 11 are learned most quickly and significantly better than the others. In ordering the remaining three structures in each of these two groups, Structures 6 and 14 may be assigned the lowest rank because of the fewer total number of relations appearing in them (only four), while the ordering of 4 and 12, and 5 and 13 could only be assigned to a more powerful influence of common origin than common goal.

Considering the structures in Group IV from this point of view, it is quite reasonable for Structure 10 to show the slowest learning rate (having none of our relations embedded in it). However, the poor showing of Structure 7 does not simply fit this schema. Obviously, much further evidence will have to be adduced—probably with structures involving more than three elements—in order to arrive at a metric of structural relations.

The appearance of probability matching in a task ostensibly designed for other purposes we find most interesting. The $S$s, given an event matching task of moderate difficulty, respond to the probability structure of responses surprisingly quickly (cf. Grant, Hake, & Hornseth, 1951) and without necessary reference to the event structure. To what extent this behavior is a function of hypothesis formation in the $S$s is difficult to estimate. The parallel between probability matching in our situation, with its low payoff, and probability matching in a two-choice situation (as against maximizing) seems obvious and deserves further examination.

Finally, it is of note that probability matching apparently is most marked when event matching proceeds most slowly. In the comparison between Groups III and V, for example, the fact that Group III learns more quickly may be due to the stress in the instructions on the detection of checkmarks; however, probability matching cannot easily

TABLE 3

NUMBER AND TYPE OF RELATIONS IN DIFFERENT STRUCTURES

| Relations | Structures | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 3 & 11 | 4 & 12 | 5 & 13 | 6 & 14 | 7 | 8 | 9 | 10 |
| Symmetry | 3 | 1 | 1 | 1 | 2 | 2 | 0 | 0 |
| Transitivity | 0 | 1 | 1 | 1 | 0 | 0 | 2 | 0 |
| Common origin | 1 | 3 | 1 | 1 | 2 | 0 | 2 | 0 |
| Common goal | 1 | 1 | 3 | 1 | 0 | 2 | 2 | 0 |

be assigned to this variable. The general proposition that as event discrimination becomes easier, rate of probability matching decreases and vice versa, deserves exploration.

## SUMMARY

Sixteen groups of 12 Ss each were required to learn the presence or absence of a checkmark associated with six pairs of nonsense syllables constructed out of three CVC syllables. The 16 groups differed in the structural relations among pairs of syllables and the presence or absence of checkmarks. The data were analyzed in respect to (a) the number of checkmarks present in any one structure, and (b) the presence of certain logical and quasilogical relations among the three syllables.

The results showed: (a) Number of checkmarks (or noncheckmarks) predicted the acquisition of structures, based apparently on discriminability of the checkmark-noncheckmark dichotomy. (b) Within three groups of structures with the same number of checkmarks or noncheckmarks, logical structure showed significant effects on acquisition. (c) The probability structure of the required responses (checkmarks) showed a striking effect on Ss' behavior. The Ss exhibited probability matching, i.e., emission of the required percentage of responses, in the absence of event matching, i.e., correct response to the stimulus information.

## REFERENCES

DeSoto, C. B. Learning a social structure. *J. abnorm. soc. Psychol.*, 1960, **60**, 417–721.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with different percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Harary, F., & Norman, R. Z. *Graph theory as a mathematical model in social science.* Ann Arbor: University of Michigan Institute for Social Research, 1953.

# SOME PROPERTIES OF SACCHARIN AS A REINFORCER [1]

GEORGE COLLIER

*University of Missouri*

Saccharin appears to have only one property in common with the sugars: it tastes sweet. It differs from sucrose both in its other biological properties and in its physical properties. Saccharin differs from sucrose in that it is not a sugar, it is nonnutritive (75–90% eliminated within 24 hr. in urine and the remainder in the feces, Carlson, Eldridge, Martin, & Foran, 1923), it is hypotonic[2] at all concentrations examined, Table 1 (approximately 3.7% by weight would be isotonic), and it has a substantially lower preference threshold (of the order of .01%, Stellar, 1960). Saccharin elicits salivary, gastric, and intestinal secretions, but to a lesser degree than sugar. It decreases intestinal absorption of water in proportion to its concentration by some nonosmotic mechanism. It appears in the blood, lymph, cerebrospinal fluid, tears, and mammary secretions following its ingestion in proportion to its concentration (Carlson et al., 1923). Saccharin intake does not cause a compensatory decrease in food intake on an ad lib. feeding schedule as does sugar (Hausmann, 1933), nor does it affect weight, mortality, state of organs (Fitzhugh, Nelson, & Frawley, 1951), with the possible exception that at high concentrations it may result in a slight weight loss when fed in daily diet

(Thompson & Mayer, 1959). These differences and similarities provide a means of examining the possible mechanisms involved in food reinforcement. The present study explores saccharin as a reinforcer, comparing the functions obtained with those obtained from sucrose.

## METHOD

*Apparatus.*—Eight Skinner boxes (Collier & Myers, 1961) delivering liquid reinforcements were used. The solutions were prepared from soluble saccharin (Merck).

*Subjects.*—The 20 *S*s of Exp. 1, the 24 *S*s of Exp. 3, and the 12 *S*s of Exp. 4 were 120–150-day-old female rats. They had been used in a previous experiment in the same apparatus with sucrose solutions. Equal numbers of *S*s from each of the three concentrations in the preceding experiment were assigned to the conditions in these experiments. They were maintained on a 23-hr. food privation schedule.

The 64 *S*s of Exp. 2 were 90-day-old naive female rats. They were maintained on a 10-gm. per 24-hr. food privation schedule. The daily ration was placed in *S*'s cage immediately after running.

The 16 *S*s of Exp. 5 were 180-day-old male rats. They had been used previously in a water reinforcement experiment. In the "thirsty" portion of the experiment *S*s were maintained on 1 hr. water, immediately following a session, and free food. In the "hungry" portion of the experiment *S*s were

[2] The extracellular fluid concentration is of the order of 310 milliosmols for the rat.

TABLE 1

OSMOLARITY OF SACCHARIN CONCENTRATIONS USED

| % Concentration | Osmolarity[a] (Milliosmols) | Calculated Values |
|---|---|---|
| .1 | 8 | 8.2 |
| .3 | 27 | 24.8 |
| .9 | 77 | 75.4 |
| 2.7 | 225 | 229.6 |

[a] These values were obtained by freezing point determinations.

maintained on a 10-gm. per 24-hr. schedule and free water.

All rats were of the Sprague-Dawley strain (Holtzman Company), and maintained on Purina chow and tap water.

*Procedure.*—In Exp. 1, four groups of 5 Ss were run, one at each of four concentrations (.1, .3, .9, and 2.7%) on a 1-min. FI schedule for nine 30-min. sessions. Each reinforcement delivered .03 ml. Following a 3-day break, the same groups were run at the same concentrations for 7 days on a 4-min. FI schedule.

Experiment 2 is a partial replication of Exp. 1 with order controlled. Four concentrations (.1, .3, .9, and 2.7%) were combined factorially with two fixed intervals between reinforcements, and two orders of presentation of the intervals (1 min.-4 min. and 4 min.-1 min.). Each S underwent both intervals, spending 6 consecutive days on each. A reinforcement delivered .1 ml. of solution. Sessions were of 20 min. duration. Two replications of 32 rats each were run. All Ss trained for 8 days on the saccharin concentration used subsequently. Some difficulty was experienced on training Ss to respond to the magazine at the higher concentrations. Those Ss lost were replaced.

In Exp. 3 two intervals, 1 and 4 min., and 3 concentrations (.1, .3, and .9%) were combined factorially with 4 Ss assigned to each combination. The first 9 days of BP were for .1 ml. of saccharin solution per reinforcement, the next 5 days of BP were for .3 ml. per reinforcement, and the final 9 days for .03 ml. per reinforcement. Sessions were 30 min. in duration.

In Exp. 4 three groups of 4 Ss each were run, one at each of the three concentrations, .3, .9, and 2.7%. A 1-min. FI schedule was used. The sessions were 1 hr. in length. In the first six sessions, .03 ml. reinforcement was used, followed by a 4-day break, 4 days on .1 ml. reinforcement, a 4-day break, and then finally 4 days on .3 ml. reinforcement.

Experiment 5 is a partial replication of Exp. 4 with the effect of order controlled. Two concentrations (.1 and 2.5%) were combined factorially with four volumes per reinforcement (.04, .08, .16, and .32 ml.) and four orders of presentation of the volumes. Each S spent 8 consecutive days at each volume. A 1-min. FI schedule and a 20-min. session were used. During the first cycle of 32 days, Ss were thirsty. During the second cycle of 32 days, Ss were hungry.

In Exp. 6, four concentrations (.1, .3, .9, and 2.7% saccharin) were combined factorially with two levels of deprivation. The high deprivation group received 8 gm. of Purina

chow and the low deprivation group 16 gm. immediately following the experimental session. A 1-min. FI schedule was used for the eight 20-min. sessions.

## RESULTS

The main results of Exp. 1 are presented in Fig. 1. The total number of bar pressing responses (BP) averaged over the last 2 days on each interval is plotted against log concentration for each FI interval. Analysis of variance of these data for Concentration, Interval, and Sessions showed the main effects of Concentration and Interval significant ($P \leq .01$, $F_C = 6.12$, $df = 3/16$; $P \leq .01$, $F_I = 49.55$, $df = 1/16$). On the final day of the 1-min. and 4-min. conditions 97% of the rewards possible were received and 100% of those received were consumed. Thus, all 1-min. groups and all 4-min. groups consumed equal volumes of solution. When the total BPs per session are compared with those from the previous sucrose experiment in which Ss served, it is apparent that there was a large, significant decline from the rate for comparable sugar solutions
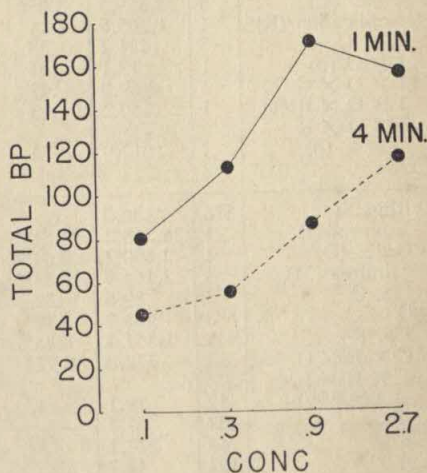


FIG. 1. The number of BPs/session as a function of concentration and interreinforcement interval in Exp. 1.
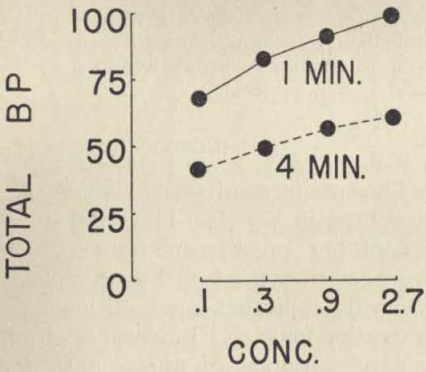
FIG. 2. The number of BPs as a function of concentration and interval in Exp. 2.

to a lower terminal rate for saccharin. Experiment 1, using .03 ml. per reinforcement, shows that rate of BP was an increasing function of concentration up to at least 2.7%, nine times the highest reported free ingestion preference value, on a 4-min. FI schedule and an increasing function

of concentration up to at least .9% on a 1-min. FI schedule.

The basic results of Exp. 2 are presented in Fig. 2. An analysis of variance of these data for the average of the last 2 days of each order, grouped into 4-min. totals, is presented in Table 2. The rate of responding was an increasing function of concentration at both intervals between reinforcements. The rate of responding within a session declined significantly, the level of responding in the final 4 min. being about 68% of that of the initial 4 min. However, in Fig. 3 it is apparent that there were no differences in rate of decline over a 20-min. session.

The major results of Exp. 3 are presented in Fig. 4. Rate of responding was an increasing function over the range of concentrations examined for small volumes per reinforcement, and was an increasing then decreasing function of concentration for large volumes. At the larger volumes the highest rates were obtained for the

TABLE 2

ANALYSIS OF VARIANCE OF THE BP DATA OF EXP. 2

| Source | df | MS | F |
|---|---|---|---|
| Between $S$s | 63 | 10,936.7 | |
| Conc. (C) | 3 | 3,359.5 | 3.05* |
| Replication (R) | 1 | 1,265.6 | 1.14 |
| R × C | 3 | 421.2 | .38 |
| I × O (b) | 1 | 12.1 | .01 |
| I × O × C (b) | 3 | 469.0 | .42 |
| I × O × R (b) | 1 | 150.2 | .13 |
| I × O × R × C (b) | 3 | 617.9 | .56 |
| Error$_b$ | 48 | | |
| Within $S$s | 576 | 136.0 | |
| Interval (I) | 1 | 26,368.2 | 94.81** |
| Order (O) | 1 | 2,600.2 | 9.34** |
| Minutes (M) | 4 | 2,164.9 | 47.33** |
| I × C | 3 | 89.6 | .32 |
| O × C | 3 | 965.9 | 3.47* |
| I × R | 1 | 1,351.4 | 4.85* |
| C × R × O | 3 | 772.8 | 2.77* |
| C × R × I × O × M (b) | 12 | 19.1 | .51 |
| Error$_w$ | 432 | 67.7 | |
| $e_1$ | 48 | 278.1 | |
| $e_2$ | 192 | 45.7 | |
| $e_3$ | 192 | 37.1 | |

* $P \leq .05$.
** $P \leq .01$.

FIG. 3. The number of BPs in each 4-min. period as a function of concentration in Exp. 2.
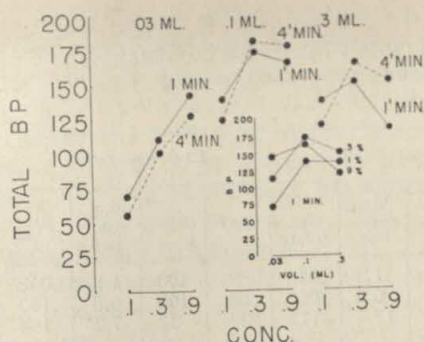
FIG. 4. The number of BPs as a function of volume per reinforcement, concentration, and interval between reinforcements in Exp. 3.

TABLE 3

ANALYSIS OF VARIANCE OF THE BP DATA OF EXP. 4

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | 11 | 325.0964 | |
| Conc. (C) | 2 | 21.02 | .05 |
| Error$_b$ | 9 | 392.67 | |
| | | | |
| Within Ss | 2148 | 8.02 | |
| Volume (V) | 2 | 836.98 | 10.12** |
| Minutes (M) | 59 | 31.30 | 5.69** |
| V × M | 118 | 5.16 | .81 |
| C × V | 4 | 292.64 | 3.54* |
| C × M | 118 | 6.46 | 1.17 |
| C × V × M | 236 | 5.46 | .86 |
| | | | |
| Error$_w$ | 1611 | | |
| e$_1$ | 18 | 82.63 | |
| e$_2$ | 531 | 5.50 | |
| e$_3$ | 1062 | 6.37 | |

* $P \leq .05$.
** $P \leq .01$.

longer intervals. The difference in level of responding between these results, for .1 ml., and those of Exp. 2 is probably the result of the fact that Ss in Exp. 2 were trained directly on saccharin.

The main results of Exp. 4 are presented in Fig. 5, in which the cumulative number of responses per minute for each combination of volume and concentration averaged over the last two sessions in each cycle is plotted against time. An analysis of variance over Concentration, Volume, and Minutes is presented in Table 3.

An examination of Fig. 5 shows that at small volumes rate of BP is an increasing function of concentration, at large volumes it is a decreasing function, and that the rate of decline in responding is a function of concentration but independent of the volume.
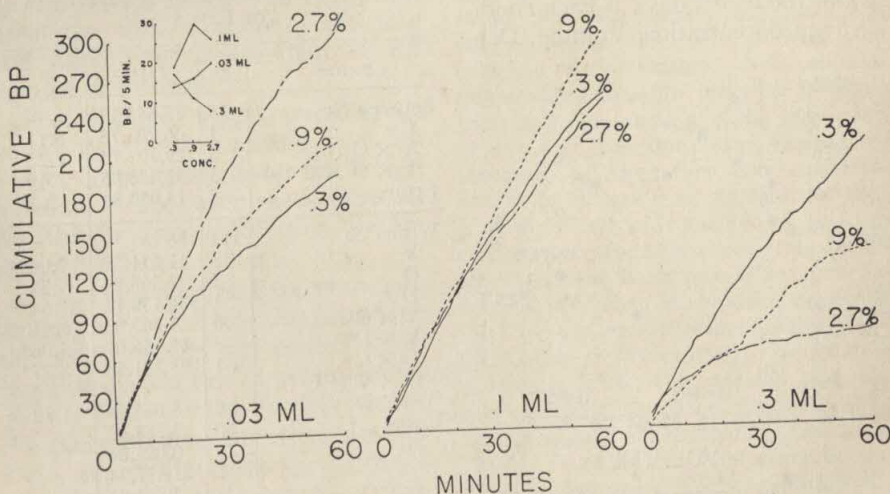


FIG. 5. The cumulative number of BPs and the number in the first 5 min. as a function of volume and concentration of reinforcement in Exp. 4.

## TABLE 4

NUMBER OF REINFORCEMENTS RECEIVED (#R), PERCENTAGE CONSUMED (%),
AND VOLUME OF SOLUTION AND AMOUNT OF SOLUTE CONSUMED
IN EXP. 4

| Vol. (ml.) | .3 Concentration | | | | .9 Concentration | | | | 2.7 Concentration | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | #R[a] | %[b] | Vol. Solution | Gm. Solute | #R[a] | %[b] | Vol. Solution | Gm. Solute | #R[a] | %[b] | Vol. Solution | Gm. Solute |
| .03 | 58 | 100 | 1.7 | .005 | 60 | 100 | 1.8 | .016 | 59 | 100 | 1.8 | .048 |
| .1 | 52 | 100 | 5.2 | .016 | 60 | 100 | 6.0 | .054 | 60 | 100 | 6.0 | .162 |
| .3 | 56 | 99 | 16.6 | .048 | 44 | 93 | 12.4 | .111 | 33 | 72 | 7.2 | .194 |

[a] 60 possible.
[b] Percentage of reinforcements consumed when magazine operated.

This latter finding has been reported for sucrose (Collier & Myers, 1961) and for salt (Stellar, Hyman, & Samet, 1954). Total BPs per session do not show the decline observed in Exp. 1. With the longer sessions (60 min. vs. 30 min.) the decline takes place within sessions rather than between.

Table 4 shows the number of the 60 possible reinforcements received, the percentage taken, and the volume of solution and grams of solute consumed.

The major data of Exp. 5 are presented in Fig. 6. An analysis of the data of the final 2 days at each condition for Concentration, Volume, Deprivation, and Order is presented in Table 5. When Ss were hungry, rate of responding was an increasing function of volume at the low (.1%) concentration and an increasing then decreasing function of volume at the high (2.5%) concentration. When Ss were thirsty, number of responses was an increasing then decreasing function of volume at both concentrations and a decreasing function of concentration at all volumes. These volume per reinforcement curves

## TABLE 5

ANALYSIS OF VARIANCE OF THE BP DATA
OF EXP. 5

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | 15 | 23,853.52 | |
| C | 1 | 66,703.78 | 5.11 |
| V × O (b) | 3 | 29,792.73 | 2.28 |
| V × O × C (b) | 3 | 32,438.05 | 2.49 |
| Error_b | 8 | 13,050.84 | |
| Within Ss | 112 | 16,285.29 | |
| V | 3 | 47,811.59 | 7.96** |
| O | 3 | 3,268.89 | .54 |
| D | 1 | 536,130.12 | 22.64** |
| V × O (w) | 6 | 7,867.00 | 1.31 |
| V × D | 3 | 11,656.07 | 3.06* |
| D × C | 1 | 195,781.53 | 8.27* |
| O × V × C (w) × D | 6 | 5,645.44 | 1.48 |
| Error_w | 56 | 7,588.42 | |
| e_1 | 24 | 6,004.84 | |
| e_2 | 8 | 23,677.34 | |
| e_3 | 24 | 3,809.01 | |

* $P \leq .05$.
** $P \geq .01$.
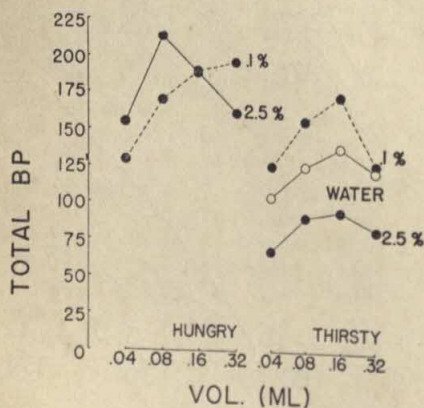


FIG. 6. The number of BPs as a function of concentration, volume per reinforcement, and kind of deprivation in Exp. 5.

for saccharin are similar to those obtained with water from a comparable group of Ss run under the same conditions (Manaster, 1962). The addition of a "small" amount of saccharin appears to lead to an increase in rate of responding above that for water while the addition of a "large" amount leads to a decrease.

Within a session, both initial rate and the rate of decline were affected by concentration and deprivation. Hungry Ss showed no noticeable decline within sessions at any combination of volume and concentration with the exception of the .04 ml.-.1% group. The differences in total number of bar presses were reflected in the initial rates, that is, the non-monotonicity of the rate-volume curve at the high concentration and the nonmonotonicity of the rate-concentration curves at the large volumes could not be attributed to postingestive effects. On the other hand, when Ss were thirsty, substantial within-session declines occurred, particularly at the high concentration. Here it is obvious that some postingestive effect was operative over the 20 min. of the session which was dependent upon the concentration of the load and apparently independent of the volume of the load.

The Ss gained weight across cycles under the thirst schedule and lost weight across cycles under the hunger schedule. The average weight for the thirsty Ss was 355 gm.; for the hungry Ss it was 289 gm.

The major data of Exp. 6 are presented in Fig. 7. It shows that both the slope of the rate-concentration function and the rate of within-session decline were affected by deprivation, the steepest slope occurring for the 8-gm. group, which averaged 199 gm. in weight, and the greatest within-session decline for the 16-gm. group which averaged 270 gm. in weight.
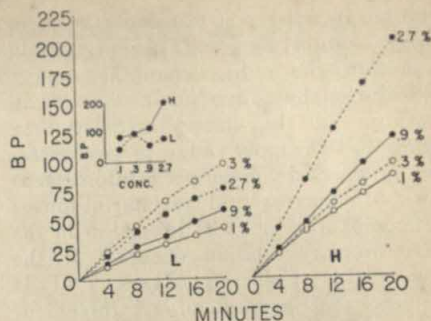


FIG. 7. The cumulative number of BPs as a function of concentration and degree of food deprivation in Exp. 6.

## DISCUSSION

Saccharin resembles sucrose in that similar functions are obtained when the parameters of reinforcement and deprivation are manipulated and the course of responding within a session examined. It is clear from these data that the supposition that there are three independent loci of the events governing the rate of responding—the proximal reinforcing stimuli, the momentary postingestive load, and the nutritive condition of the animal—is correct for saccharin as well as for sucrose.

Initial rates of responding for saccharin prove to be functions of the same dimensions of the proximal reinforcing stimulus as for sucrose, namely concentration, volume per reinforcement, and interval between reinforcements. No combination of these variables proves to be additive, both the slope and the maximum of any one function being themselves functions of the other two variables. Saccharin differs from sucrose in that it is responded to at much lower concentrations; it produces a lower maximum rate; and the increment in log concentration necessary to produce an equal increment in rate of responding is larger (cf. Collier & Myers, 1961). This latter contrasts with the parallel rate-log concentration functions obtained when similar sugars are compared (Guttman, 1954). Volume per reinforcement appears to have two effects on initial rate of responding for saccharin. In the lower range of concentrations increased

volume increases the slope of the rate-concentration function. This may merely represent the reduction of the effect of dilution by saliva or it may be an example of the classic intensity-area relation. In upper range of concentration, increased volume per reinforcement for both sucrose and saccharin lowers the point at which the inversion in the rate-intensity relation occurs. If the assumption (Collier & Myers, 1961) is correct that amount of reinforcement is an increasing function of the intensity and volume of the proximal reinforcing stimulus, then some other processes must intervene at these values. Two nonexclusive possibilities are that (a) the quality of the stimulus changes (e.g., from sweet to bitter) at the large volume high concentration combinations, and (b) that there are unconditioned withdrawal responses to intense stimuli which compete with the reinforced response. Some indication for the latter is given by the fact that the latency of the magazine response is longer at very high concentrations (Collier, 1959).

As in the preceding experiments (e.g., Collier & Myers, 1961), two sorts of within-session decrements are found, those occurring at minimal reinforcement values and those occurring for combinations of volume, interval, and concentration which result in concentrated postingestive loads. Numerous authors have attributed this latter shutoff to the increase in the osmotic pressure of the gastric load. Amount consumed as a function of concentration studies typically show a peak intake at approximately the point of isotonicity while loading studies have typically shown an increasing depression of intake as a function of hypertonicity. However, the present results show the rate of shutoff as an increasing function of concentrations, all of which were hypotonic (Table 1), and two volume-consumed studies show peak intakes at approximately .24% for a 1-hr. session (Stellar, 1960) and .44% for a 55-min. session (Cockrell, 1952), which are considerably below the point of isotonicity.[3] Thus,

[3] Interpolated from the curves presented.

we have a postingestive shutoff effect in the hungry $S$ which is proportional to concentration but is not due to the hypertonicity of the load. Similarly, for the thirsty $S$s we find a higher rate of responding and a lower rate of decline for the low concentrations than for water, and a lower rate and a faster rate of decline for the high concentration, which is again still hypotonic. These results lead to a suggestion either of some peculiar property of saccharin or of a rejection of a simple osmotic explanation of satiation. Some support for the former alternative is given by Carlson et al. (1923), who report that saccharin delays intestinal absorption by some mechanism other than the osmotic factor. The results further suggest that there are two shutoff mechanisms, one for hunger and one for thirst, which respond differentially to the same load.

The effects of nutritive condition on the rate of responding for saccharin are similar to those for sucrose (Collier & Willis, 1961). Higher rates of responding and steeper slopes of the rate-log concentration function are obtained at higher deprivations. These results, in the light of the noncaloric character of saccharin and its sustained consumption over long periods of time (e.g., Hillix, 1958), support the view that deprivation may exercise its effects on rate of responding independently of the postingestive consequences of the reinforcement. However, it should be noted that saccharin simulates sugars and other nutrient materials in some of its non-nutritive postingestive consequences, in that it elicits similar gastric reflexes, etc. and that a large part of it is absorbed before it is eliminated (Carlson et al., 1923).

When the kind (e.g., thirst vs. hunger) rather than the degree of deprivation is varied the rate-intensity relation is again affected. For thirsty $S$s the rate-intensity relation is essentially flat over the lower range of concentrations and then decreases for higher concentrations. The hypothesis that body concentration is being defended by means of an osmoreceptor at the gustatory level is not

supported in the present study since all values of saccharin used were hypotonic. This interaction of taste and kind of deprivation again suggests that deprivation may in part exercise its effect on rate of responding independently of the postingestive consequences of the reinforcement.

## SUMMARY

The relations between concentration, volume per reinforcement, interval between reinforcement, degree and kind of deprivation, and the rate of responding for saccharin solutions in the Skinner box were explored.

For hungry Ss initial rate of responding was an increasing function of concentration and volume, and a decreasing function of interval. No combination of these variables proved to be additive. The slope of the initial rate vs. log concentration function was an increasing function of deprivation while the slope of the initial rate vs. log volume function was not. When Ss were thirsty the rate vs. log concentration function became flat at the low concentrations and decreased at the high, while the rate vs. log volume function retained the same shape. Rate of shutoff, within a session, was an inverse function of the magnitude of reinforcement at low levels of reinforcement and a function of the magnitude of the load at high levels of reinforcement. The rate of shutoff did not appear to be greatly affected by deprivation.

The relations found for saccharin were similar to those found for sucrose. The implications of the differences between saccharin and sucrose, e.g., osmotic and metabolic, were examined for an account of these relations in terms of the view that there is a threefold locus of events governing the rate of responding, the proximal reinforcing stimuli, the momentary ingestive load, and the nutritive condition of the animal.

## REFERENCES

CARLSON, M. J., ELDRIDGE, C. T., MARTIN, H. P., & FORAN, F. L. Studies in the physiological action of saccharin. *J. metabol. Res.*, 1923, **3**, 451–477.

COCKRELL, J. T. Operant behavior in relation to the concentration of a nonnutritive sweet substance used as a reinforcement. Unpublished doctoral dissertation, Indiana University, 1952.

COLLIER, G. The loci of reinforcement. *Amer. Psychologist.*, 1959, **14**, 398. (Abstract)

COLLIER, G., & MYERS, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61**, 57–66.

COLLIER, G., & WILLIS, F. Deprivation and reinforcement. *J. exp. Psychol.*, 1961, **62**, 377–384.

FITZHUGH, O. G., NELSON, A. A., & FRAWLEY, J. P. A comparison of chronic toxicities of synthetic sweetening agents. *J. Amer. Pharmaceut. Ass., scient. Ed.*, 1951, **40**, 583–586.

GUTTMAN, N. Equal reinforcement values for sucrose and glucose solutions compared with equal sweetness values. *J. exp. Psychol.*, 1954, **47**, 358–361.

HAUSMANN, M. F. The behavior of albino rats in choosing foods: II. Differentiation between sugar and saccharine. *J. comp. Psychol.*, 1933, **15**, 419–428.

HILLIX, W. A. Volume ingested as a function of deprivation, taste, and nutrition. Unpublished doctoral dissertation, University of Missouri, 1958.

MANASTER, M. Volume per reinforcement as a parameter of amount of reinforcement functions. Unpublished master's thesis, University of Missouri, 1962.

STELLAR, E. Drive and motivation. In J. Field (Ed.), *Handbook of physiology.* Vol. 3. *Neurophysiology.* Washington, D. C.: American Physiological Society, 1960. Pp. 1501–1527.

STELLAR, E., HYMAN, R., & SAMET, S. Gastric factors controlling water and salt-solution drinking. *J. comp. physiol. Psychol.*, 1954, **47**, 220–226.

THOMPSON, M. M., & MAYER, J. Hypoglycemic effects of saccharin in experimental animals. *Amer. J. clin. Nutr.*, 1959, **7**, 80–85.

# ADAPTATION IN THE PERCEPTION OF VISUAL VELOCITY

## V. R. CARLSON

*National Institute of Mental Health*

The aftereffect due to prolonged viewing of a constantly moving stimulus pattern has been studied with (*a*) a stationary test pattern, (*b*) a pattern moving in the same direction slowly enough to appear stationary, and (*c*) a pattern moving in the same direction and at the same objective speed as the adapting pattern. If one views a pattern moving down for a time and then, within some number of seconds, looks at a stationary pattern, the latter will give an appearance of moving up (the so-called waterfall illusion); if the pattern is moved down at a slow rate instead of remaining stationary, the apparent upward motion can be cancelled. If the original pattern continues and a second, identically moving pattern is presented in a different part of the field, the first will appear to be moving at a slower rate than the second.

Gibson (1937, pp. 234–236; 1959, pp. 490–491) has reviewed these several effects and has pointed out that they are consistent with the proposition that adaptation involves an oppositely directed decrement in apparent velocity along the dimension of the adapting velocity. As a purely empirical generalization, one would predict that a test pattern moving in the same direction, but faster than the adapting pattern, should appear to be moving more slowly than it would in the absence of the prior adaptation. Similarly, if the test pattern is moving in the direction opposite to that of the adapting pattern and at a more or less comparable rate of speed, then the test pattern should appear to be moving faster than normal in that opposite direction. The present study is a test of these two hypotheses, utilizing a procedure for measuring perceived velocity which does not depend upon subjective awareness of the occurrence of any aftereffect.

## METHODS

*Perceptual velocity test.*—Fixating binocularly, *S* viewed an aperture (depicted in Fig. 1) through which a bright test line rotated orbitally at constant speed. Two seconds before the appearance of the test line, a tone signal was presented and at the same time the designated target was lighted. The *S* depressed a key and held it down until he judged that the test line had reached the target position beyond the point of disappearance of the test line. Response time was measured from the moment of disappearance of the test line to the instant *S* released the response key. Release of the key also extinguished the target.

This general kind of task has been used by a number of investigators in the study of tracking performance (reviewed by Brown, 1961, pp. 99–101). Gerhard (1959) and Held and White (1959) have employed it more specifically as a means of measuring perceived velocity. The present version, although similar in principle, differs in many details from the particular tests used by those investigators.

*Adaptation.*—Prior to an adaptation test trial, a continuously rotating pattern was presented in the aperture for 45 sec. The *S* made no response during the adaptation period other than to maintain fixation. The pattern was moving at a constant rate both when it appeared and disappeared, so that *S* never saw it stationary. Three identical presentations of the test line then occurred at approximately 4, 13, and 22 sec. following disappearance of the adaptation pattern. Control trials were the same as the adaptation trials except that no pattern appeared in the aperture during the adaptation period. One minute elapsed between successive test periods (whether adaptation or control).
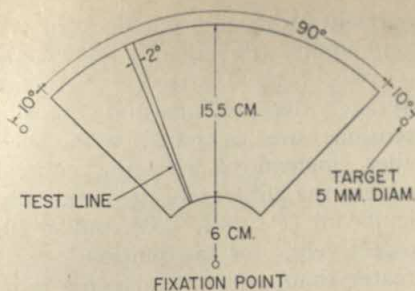
FIG. 1.   Aperture used in perceptual velocity test.

*Schedule.*—The speeds of rotation were 30°, 40°, and 53°/sec, with corresponding correct response times of .33, .25, and .19 sec., respectively.[1] At the beginning of each session $S$ performed two practice test trials at 40°/sec, one clockwise and one counterclockwise. The experimental schedule consisted of six control conditions interspersed among 18 adaptation/test conditions on each of 4 days (the nine possible speed combinations each occurred once with the adaptation and test rotations in the same direction and once in opposite directions). The order of occurrence of these conditions was carefully distributed and permuted so as to minimize possible sequential effects. Several days, usually 1 week, intervened between testing sessions.

*Apparatus.*—The $S$ was positioned 2 m. in front of a 59 × 76 cm. rear-projection, flashed opal-glass screen with the fixation point at eye level. The aperture, fixation point, and target spots were produced by cutouts in a sheet of cardboard mounted immediately behind the glass. The test line and adaptation pattern were produced by open sectors in discs mounted on a rotator with the center of rotation in line with the fixation point and $S$'s eye. The test line consisted of a 2° sector, the adaptation pattern, of 18 2° sectors spaced 18° apart. The screen, its black frame, and the fixation point were always clearly visible, but the rest of the test stimuli appeared only at the appropriate times during an actual trial.

*Subjects.*—Fifteen male and 4 female junior college students were paid to serve as $S$s in the main experiment. The data for the females fell within the distributions for

[1] All designations of measure in degrees in this paper refer to angular distance around the fixation point in the frontal plane of the fixation point, not to visual angle.

the males and showed no consistent tendencies toward deviation from the data of the males. The two groups were therefore treated as a single group of 19 $S$s.

*Preliminary work.*—A variety of $S$s have shown very high correlations between response time in this perceptual velocity test and presented speed of rotation of the test line, although absolute response time has varied appreciably from one individual to another. No relationship was found between response time in this task and simple reaction time to the cessation of a light, in agreement with Held and White (1959).

In a preliminary adaptation experiment with 11 $S$s (other than those used in the main experiment) only control trials were presented on a first day, only adaptation trials on a second day, and only control trials again on a third day. Overall response-time level tended to shift toward generally increased response times after the first day, and it was for this reason that the control trials were distributed among the adaptation trials in the main experiment. The results of the preliminary experiment were otherwise essentially the same as those reported below.

## RESULTS

Clockwise vs. counterclockwise rotation, as such, had no differential effect, so these conditions have been combined. Rotations in the Same direction thus refer to conditions in which both the adaptation pattern and the subsequent test line turned clockwise or both counterclockwise; rotations in Opposite directions refer to those conditions in which one turned clockwise and the other counterclockwise.

An overall test for effect on the first test-line presentation following each adaptation period was made according to whether speed of the adaptation pattern was greater than, equal to, or less than the speed of the test line and whether rotations were in the Same or Opposite directions (Table 1). Since each of these six categories combines three different speed conditions (the appropriate combinations of 30°, 40°, and 53°/

TABLE 1

AVERAGE DIFFERENCE IN PERCENTAGE
RESPONSE-TIME ERROR (CONTROL
MINUS ADAPTATION) FOR FIRST
TRIAL FOLLOWING
ADAPTATION

| Adaptation vs. Test Direction | Adaptation Speed Relative to Test Speed | | |
|---|---|---|---|
| | Greater | Equal | Less |
| Same | | | |
| Mean | −10.5 | −14.7 | 1.6 |
| SD | 9.0 | 23.3 | 20.9 |
| P[a] | <.001 | <.02 | ns |
| Opposite | | | |
| Mean | −0.6 | −4.4 | −8.7 |
| SD | 6.4 | 11.8 | 18.4 |
| P | ns | <.10 | <.05 |

[a] Probability associated with the $t$ value for the difference of the mean from zero, $df = 18$.

by itself, although a tendency in the nonpredicted direction is suggested for two of these categories. If the average values for Same and Opposite rotations are compared with each other, instead of with the control values, the differences are generally significant ($P < .001$, $< .05$, and $< .10$, respectively, for adaptation speed greater than, equal to, and less than test speed).

More detailed results for rotations in the Same direction are presented in Fig. 2. The corresponding curves for rotations in Opposite directions are not shown, since they were not, in any specific instance, significantly different from the control values. Each adaptation period was followed

sec), response-time error expressed as a percentage of the correct response time for each given test speed was used as the score. According to the empirical hypothesis under test, the control-minus-adaptation differences for rotations in the Same direction should be negative (relatively longer response time indicating slower apparent speed following adaptation as compared to control); the differences for rotations in Opposite directions should be positive (test line appearing to move relatively faster following adaptation to a pattern moving in the opposite direction). For only two of the six categories in the table do the results agree with such a prediction. These two (adaptation speed equal to, or greater than, test speed, rotations in Same direction) represent consistent effects, each of the three conditions within each category at least approaching a significant difference ($P < .10$) when tested separately. No single within-category condition for rotations in Opposite directions reached a significant effect



FIG. 2. Response-time errors on three successive test trials following adaptation (test and adaptation rotations in the Same direction).

by three successive, identical test trials, occurring at approximately 4, 13, and 22 sec. after disappearance of the adaptation pattern. A lower value on the ordinate represents longer response time, indicating relatively slower perceived speed of the test line.

The three control points for test trials at the middle speed, 40°/sec, do not differ from each other within the limits of error. For both 30°/sec and 53°/sec, the second trial tends to be different from the first ($P < .05$), effective speed of the test line shifting toward a relatively slower value for 53°/sec and toward a relatively faster value for 30°/sec. These trends are generally discernible in the curves for the adaptation trials as well.

The clearest aftereffect is a displacement of the points downward when the velocity of the adaptation pattern was greater than the velocity of the test line. A lesser effect, but in the same direction, occurred when the adaptation and test-line velocities were equal. In all of the present data the only instance of a significant increase in perceived speed due to prior adaptation was produced by adaptation at 30°/sec on test trials at 53°/sec rotations in the Same direction. All three points for this curve are significantly higher than the corresponding control points ($P < .01$).

Since the preliminary experiment suggested an overall shift toward longer response times with succeeding days, the data were also analyzed for such a trend. The mean response times for the control trials were .10, .06, and .04 sec. longer for speeds 30°, 40°, and 53°/sec, respectively, on the fourth as compared to the first day. These increases were progressive over the 4 days, but there was no consistent trend from beginning to end of testing sessions within days.

## Discussion

Motion aftereffects of the waterfall type have represented a class of effects in which a decrement in effective velocity accrues to a test stimulus moving in the same direction as the adapting stimulus and at the same, or a slower, speed. The significant negative aftereffects which occurred in the present experiment fall within this class. The procedure used here is incapable of utilizing a stationary test stimulus, but Johansson (1956) has devised a similarly nonsubjective technique representing the stationary test condition. His results also indicate a negative aftereffect of motion. The more objective procedure, therefore, appears to produce results which are consistent with previous subjective determinations of motion aftereffect, at least with respect to direction.

Smith and Sherlock (1957) have pointed out that when a pattern moves through an aperture the apparent frequency of passage of contours is confounded with apparent speed. Such confounding was the case here with respect to the adaptation pattern and presumably contributed to the impression of its speed. Also, the time the test line was visible was confounded with its speed through the aperture, and it is possible that time-judgment is an important aspect of this type of task (Gerhard, 1959). But if the adaptation effects were due solely to the factors of frequency or time, then the direction of rotation of the test line relative to that of the adaptation pattern should not have made any difference. Since this relation was of major importance in the results, we can conclude, tentatively at least, that there was adaptation to velocity, not just to speed or a nondirectional component such as frequency or time. One aspect of the results, however, does suggest some nondirectional adaptation. That was the tendency for the three successive test trials in a set, whether following the adaptation

pattern or not, to shift toward a higher value for the slowest speed and toward a lower value for the fastest speed. This effect would be consistent with the notion that an Adaptation Level tended to become established at or near the middle speed (or response time), dependent upon a cumulative effect of all preceding trials.

One of the directional effects differentially supports Adaptation Level theory rather than the principle of negative aftereffect: Adaptation at 30°/sec produced an apparent increase upon the test speed at 53°/sec, when both rotations were in the same direction. However, the two other comparable conditions, 30° upon 40° and 40° upon 53°/sec, did not. Moreover, the lack of negative aftereffect with rotations in opposite directions also appears inconsistent with Adaptation Level theory. According to Helson (1959), ". . . adaptation has not only negative aftereffect but positive and negative effects simultaneously: high AL sensitizes to negative qualities, low AL to positive qualities, and intermediate AL to both positive and negative qualities . . ." (pp. 572–573).

In any case, it is specifically with respect to Gibson's generalization of negative aftereffect that the present findings do not agree. Gibson (1959) conceives of a single dimension as consisting of values varying from high in one direction through zero to high in the opposite direction. Adaptation to any particular value along the entire dimension reduces the difference between that value and zero and simultaneously shifts all values along the dimension in the same algebraic direction (p. 490). If the implication of the present findings is correct, however, the principle would appear not to apply to perceived velocity, whether one understands the pertinent variable of stimulation to be angular optical motion across the retina or the "shear" relation between a pattern and the edges of an aperture (Gibson, 1958, p. 168).

On the other hand, the existence of ganglion cells in the vertebrate retina which are differentially sensitive to the direction of motion of a contour (Hubel & Wiesel, 1959, pp. 581–584; Maturana, Lettvin, McCulloch, & Pitts, 1960, p. 159) suggests a possible basis for a lack of aftereffect when the adapting and test motions are in opposite directions. As far as such cells, or processes dependent upon them, would be concerned, the effects of stimulation in one direction would be expected to have little or no effect upon subsequent stimulation in a different direction. This is a highly speculative hypothesis at this point, but it seems likely to require consideration in future theorizing about motion perception.

## SUMMARY

The effect of adaptation to an orbitally rotating pattern on a subsequently presented moving test stimulus was assessed using a procedure in which S is unaware of the occurrence of aftereffect. When adaptation and test motions were in the same direction, results were generally consistent with already known aftereffects of the waterfall-illusion type. But little or no aftereffect occurred when adaptation and test motions were in opposite directions. This finding agrees neither with Adaptation Level theory nor with Gibson's principle of negative aftereffect. It may, however, be related to the recent discovery of retinal units which are differentially sensitive to the direction of stimulus movement.

## REFERENCES

Brown, R. H. Visual sensitivity to differences in velocity. *Psychol. Bull.*, 1961, 58, 89–103.

Gerhard, D. J. The judgment of velocity and prediction of motion. *Ergonomics*, 1959, 2, 287–304.

Gibson, J. J. Adaptation with negative after-effect. *Psychol. Rev.*, 1937, 44, 222–244.

Gibson, J. J. Research on the visual perception of motion and change. In, *Proceedings second symposium on physiological psychology*. Washington, D. C.: Office of Naval Research, (ONR symp. Rep. No. ACR-30) 1958. Pp. 165–176.

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

HELD, R., & WHITE, B. Sensory deprivation and visual speed: An analysis. *Science*, 1959, **130**, 860–861.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 565–621.

HUBEL, D. H., & WIESEL, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, 1959, **148**, 574–591.

JOHANSSON, G. The velocity of the motion after-effect. *Acta Psychol., Amst.*, 1956, **12**, 19–24.

MATURANA, H. R., LETTVIN, J. Y., McCULLOCH, W. S., & PITTS, W. H. Anatomy and physiology of vision in the frog (*Rana pipiens*). *J. gen. Physiol.*, 1960, **43**, 129–171.

SMITH, O. W., & SHERLOCK, L. A new explanation of the velocity-transposition phenomenon. *Amer. J. Psychol.*, 1957, **70**, 102–105.

# SUPPLEMENTARY REPORT: EFFECT OF MODE OF RESPONSE ON JUDGMENT OF FAMILIAR SIZE [1]

A. V. CHURCHILL

*Defence Research Medical Laboratories, Toronto, Canada*

Recent studies have compared the accuracy of estimation of the size of familiar objects from memory, with the accuracy of estimations obtained when the same objects were presented visually. Bolles and Bailey (1956) had Ss give verbal estimates of the size of 54 familiar objects from memory, followed by verbal estimates of the size of the same objects when the objects were visible. The "familiar" objects were present in the immediate environment and included items ranging from pencils and books to furniture and automobiles. The procedure suggests that the objects were familiar in terms of the recency of their appearance in S's environment.

McKennell (1960) had Ss draw lines to represent their estimates, from memory, of the size of nine common objects, followed by estimates of the size of the same objects when they were presented visually. The "familiar" objects in this study included such standard-sized items as a 9-in. rule and a cigarette, and such non-standard-sized items as a medicine bottle and a penknife.

The present study was designed to examine the reliability of the results obtained under the two response conditions, and to determine the effect of the mode of response, when estimating the size of familiar objects from memory.

*Method.*—Six laboratory personnel served as Ss. The objects to be estimated were verbally identifiable as being of a specific size, and were familiar to Ss through daily exposure in the environment. The group of objects provided a series of 40 dimensions, ranging from $\frac{5}{16}$ in., the diameter of a government issue pencil, to 12 in., the length of a Province of Ontario automobile license plate.

The S sat at a table in front of a screen which eliminated his visual reference to objects in the room which might have given cues to the size of the dimension being estimated. Under the physical-response condition, S was required to separate two straight-edges, which were mounted on tracks over a continuous strip of paper, and to mark off his estimate with a nonstandard pencil. Each marked-off estimate was removed from

S's view before continuing with the series. Under the verbal-response condition, S sat in the same position, with his hands folded to eliminate any tendency towards making physical estimates between his hands, and was asked to give a verbal estimate of the specified dimension. The physical-response condition preceded or followed the verbal-response condition for alternate Ss. The entire procedure was repeated in one session, resulting in four sets of estimates of the 40 dimensions by each S, two physical and two verbal. Each series of 40 dimensions was presented in random order.

*Results.*—Product-moment correlations were calculated, for each S, between the 40 measured dimensions and each of the four sets of estimates. Correlations were calculated between the two sets of physical estimates, between the two sets of verbal estimates, and between the means of the two sets of physical estimates and the means of the two sets of verbal estimates, for each S. The resultant 42 correlation coefficients ranged from .92 to .99.

The reproducibility of the physical responses was shown by correlations of .98 to .99 between the first and second trials, and of the verbal responses by correlations of .93 to .99. The relatively small mode-of-response effect was shown by correlations of .96 to .99 between the means of the two sets of physical responses and the means of the two sets of verbal responses.

These results are consistent with those reported by Bolles and Bailey (1956), in that Ss achieved a high degree of accuracy when making verbal estimates of the size of familiar objects on the basis of memory alone, and with those of McKennel (1960), in that Ss achieved a high degree of accuracy when making physical estimates of the size of familiar objects on the basis of memory alone.

The correlations between the estimates obtained under the two response conditions suggest that the mode of response had little effect on the accuracy of estimation of the size of familiar objects when the estimates were made from memory. The tendency towards slightly higher correlations under the physical-response condition was probably

due to the fact that Ss tend to verbalize their estimations on a discrete scale, i.e., $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, or 1 in., which would not apply to physical estimations on a continuous scale. This rounding-off effect, under the verbal-response condition, increased with the increase in the dimension to be estimated, thus, estimates of dimensions between 1 and 6 in. showed 30% to be rounded off to an even inch, whereas, estimates of dimensions between 6 and 12 in. showed over 80% rounded off to an even inch.

That Ss apparently did not remember the numerical value of verbal estimates from trial to trial is suggested by the fact that only 37% of the pairs of verbal estimates differed by $\frac{1}{16}$ in. or less, while 30% of the pairs of physical estimates differed by $\frac{1}{16}$ in. or less.

Relative error, i.e., error as a percentage of the dimension estimated, was 10% or greater in 64% of the verbal estimates, and in 57% of the physical estimates. All estimates of dimensions less than $1\frac{1}{2}$-in. dimension were less than 1 in.; approximately 50% of the $1\frac{1}{2}$-in. dimension were less than 1 in.; no estimate of dimensions of 1 in. or greater was less than 1 in.

Estimates of some of the dimensions suggested that Ss overestimate familiar objects remembered on a small background more than when these same objects are remembered on a larger background, i.e., a license plate on a small European automobile is overestimated more than one remembered on a standard American automobile.

## REFERENCES

BOLLES, R. C., & BAILEY, D. E. Importance of object recognition in size constancy. *J. exp. Psychol.*, 1956, 51, 222–225.

McKENNELL, A. C. Visual size and familiar size: Individual differences. *Brit. J. Psychol.*, 1960, 51, 27–35.

(Received June 9, 1961)

---

# SUPPLEMENTARY REPORT: AN EXAMINATION OF AN ASPECT OF THE GELB EFFECT [1]

JACOB BECK [2]

*University of Pennsylvania*

In his experiment with a hidden light source, Gelb (1929) reported that the introduction of a small bit of white paper in front of a spinning black disk changed the appearance of the disk from a faintly illuminated white to a brilliantly illuminated black. Recently, Stewart (1959) showed that this change in disk lightness is consistent with the usual laws governing lightness contrast. Beck (1961) suggested that the corresponding change from a faintly to a brilliantly illuminated field is the consequence of the increased luminance of the area now seen as white. He (Beck, 1959, 1961) reported that the judgment of illumination of a visual field consisting of discriminable areas of differing but uniform luminances is strongly influenced by the luminance of the highest reflecting area, i.e., the brightness of the area seen as white. The present experiment tested this suggestion by obtaining judgments of lightness and illumination for a situation which was in principle the same as that of Gelb's.

*Method.*—Beck's (1961) apparatus and procedure were used. The Os adjusted the illumination on a comparison surface until it appeared equal to that of the standard while viewing each with monocular vision and a motionless head in a completely dark room. The standard and comparison surfaces were 8 × 13 in. white, gray, and black smooth matte papers. The reflectances of the papers were, respectively, 84%, 21%, and 7%. The standard surface was half white and half black. Two comparison surfaces were used: One was composed of two luminance levels, equal sections of gray and black; the second was composed of three luminance levels, equal sections of white, gray, and black.

All luminance measurements were taken with a Spectra brightness spot meter. The experimental conditions and data are presented in terms of the incident illumination values which were computed from the luminance measures and the reflectances of the papers. The incident illumination on the standard surface was 1.79 ft-c and corresponded to a luminance of the white area equal to 1.5 ft-L. Thus, to equate the maximum luminances of the comparison surfaces to that of the standard, Os should

TABLE 1

MEDIANS OF Os' LIGHTNESS MATCHES

| Match | Standard Surface | | Two-Level Comp. Surface | | Three-Level Comp. Surface | |
|---|---|---|---|---|---|---|
| | Median | $Q_1$–$Q_3$ | Median | $Q_1$–$Q_3$ | Median | $Q_1$–$Q_3$ |
| White | 63% | 63%–63% | 63% | 43%–63% | 63% | 63%–63% |
| Gray | | | | | 25% | 25%–31% |
| Black | 12% | 8%–12% | 22% | 19%–28% | 12% | 8%–16% |

have adjusted the incident illumination on the two-level comparison surface to 7.14 ft-c and on the three-level comparison surface to 1.79 ft-c. For each comparison surface, O made 10 separate illumination matches with E alternately setting the incident illumination to a point either too high or too low. The order in which the two comparison surfaces were presented was alternated. At the conclusion of the illumination matches for each comparison surface, the illumination on the standard surface remained at the value used in the experiment while the illumination on the comparison surface was set at the median of O's matches. The Os were then asked to match the lightnesses of the standard and comparison surfaces with a chart of Hering grays. Ten Os were used. All had normal vision and were naive regarding the experiment.

*Results and discussion.*—Table 1 shows the medians and interquartile ranges of the reflectances of the samples on the Hering chart matched to the papers composing the standard and comparison surfaces. In agreement with Gelb's (1929) finding, the table indicates that the area of maximum reflectance on each surface was seen as light gray or white and the other areas were seen as gray or black depending upon their relative reflectances. On the two-level comparison surface, the medians of Os' lightness matches of the gray and black papers were 63% and 22%. However, due to the white paper on the three-level surface, the gray was now matched to a sample reflecting 25% and the black to a sample reflecting 12%.

Corresponding to the darkening of the papers on the three-level surface, the illumination matches were lower. On the two-level surface, the median illumination match was 8.57 ft-c with an interquartile range of 7.62–9.52 ft-c while on the three-level surface the median illumination match was 2.38 ft-c with an interquartile range of 1.79–2.74 ft-c. However, the luminances of the maximum reflecting areas on the two- and three-level comparison surfaces corresponding to the median illumination settings were similar, 1.8 ft-L and 2 ft-L. This is consistent with

the previous findings that Os' illumination judgments of a visual field consisting of surfaces of different lightnesses (and placed so that neither shadows nor highlights are present) are greatly influenced by the brightness of the area seen as white. On both the two- and three-level comparison surfaces, the maximum luminances corresponding to Os' median illumination matches were higher than the maximum luminance, 1.5 ft-L, of the standard surface. In part, this may be the result of an enhanced impression of illumination of the standard surface due to the juxaposition on it of areas of great lightness difference (Beck 1959).

Katz (1935, p. 283) hypothesized that the judgment of illumination is based on the insistence of the field, i.e., the average luminance of the field. The average luminances of the standard and comparison fields are equated when the incident illumination on the two-level comparison surface is set at 5.71 ft-c and on the three-level comparison surface at 2.14 ft-c. On the two-level comparison surface, Os' illumination matches are closer to the setting which equates the maximum rather than the average luminances of the standard and comparison surfaces. On the three-level comparison surface, the settings equating the average and maximum luminances are too similar to be distinguished. However, the earlier experiments by Beck (1959, 1961) showed that for the simplified visual field here considered Os' judgments of illumination are more strongly influenced by the maximum luminance reflected from the field than by the average luminance.

REFERENCES

BECK, J. Stimulus correlates for the judged illumination of a surface. *J. exp. Psychol.*, 1959, 58, 267–274.
BECK, J. Judgments of surface illumination and lightness. *J. exp. Psychol.*, 1961, 61, 368–377.
GELB, A. Die "Farbenkonstanz" der Sehdinge. In A. Bethe, G. v. Bergmann, G. Embden, & A. Ellinger (Eds.), *Handbuch der normalen und pathologischen Physiologie*. Vol. 12, Part I. Berlin: Julius Springer, 1929. Pp. 594–678.
KATZ, D. *The world of color.* London: Kegan Paul, 1935.
STEWART, E. C. The Gelb effect. *J. exp. Psychol.*, 1959, 57, 235–242.

# ON THE INHIBITORY EFFECTS OF A SECOND STIMULUS FOLLOWING THE PRIMARY STIMULUS TO REACT

HARRY HELSON AND JOSEPH A. STEGER

*Kansas State University*

Many conditions affecting simple spot reactions to visual, auditory, and tactile stimuli have been investigated (cf. Teichner, 1954) but, so far as we have been able to ascertain, the following phenomenon has not been hitherto reported in the literature: if a second stimulus *follows* the primary stimulus to react, reaction time (RT) is *lengthened* compared with RT when only a single stimulus is employed. Todd (1912) found that RT to a signal (visual, auditory, tactile) was *lengthened* when one or two other stimuli *preceded* the stimulus to react. The findings reported here also appear to be unique in that the intervals after onset of the primary stimulus ($S_1$) during which the second stimulus ($S_2$) still exerts a significant influence on RT are very long, considering the magnitude of RT to a single stimulus. In this study we shall be concerned with the case where $S_1$ and $S_2$ are both in the same sense modality (vision), leaving the report of an investigation of heteromodal effects to a later publication.[1]

[1] The phenomenon discussed in this article was discovered by the senior author in 1925 when attempting to condition sensory

## METHOD

*Subjects.*—The $Ss$ were 15 male students enrolled in an introductory psychology course ranging from 18 to 21 yr. of age. They were randomly divided into two groups, Group E (experimental), consisting of 10 $Ss$, and Group C (control), consisting of 5 $Ss$.

*Apparatus and procedure.*—The apparatus consisted of a black panel board on which were mounted three small neon lamps (GE 51) spaced $\frac{3}{4}$ in. apart horizontally with a small red fixation light slightly above the center lamp. A response button was conveniently placed for $S$ to press as soon as the primary light ($S_1$) went on. The $E$ sat behind the panel board and by pressing a single button actuated $S_1$, the Standard Electric clock which measured the reaction times in .01 sec., and the Hunter interval timer which operated the current on the second stimulus ($S_2$) at intervals ranging from 10 to 180 msec. in steps of 10 msec. Both $S_1$ and $S_2$ stayed on until $S$ responded since it is known that duration of a flashing visual stimulus affects

responses. In order to conceal the purpose of the earlier experiment from $Ss$ they were required to react to a light (tone) which was followed by a tone (light) 75 msec. later in most of the trials. In some trials the second stimulus was omitted and it was then noticed that RTs were shorter when only a single stimulus was presented than when the second stimulus was given. We have now completed an investigation with light and tone as the stimuli with results similar to those reported here for unimodal stimulation.

RT. The room in which the experiment was conducted was dimly lighted and shielded from external sounds. The luminances of the stimuli were not determined but they were distinctly visible against the black panel in the dim light and were not noticeably different from one another.

The three visual stimuli were employed for the purpose of counterbalancing position of $S_1$ vis-à-vis $S_2$, the interpolated stimulus. $S_1$ was thus either on the right or to the left of the middle stimulus, the latter always serving as $S_2$. The actual procedure can best be understood from the instructions given to $Ss$:

This is an experiment in simple reaction time. You are to respond to a light as quickly as you possibly can by pressing the button under your index finger.

Note the lights facing you. I will tell you which end light you are to watch. If I say "left," you will watch the light on your left and react to it as quickly as possible. If I say "right," you will watch the end light to your right and react to it as quickly as possible.

The procedure will be this: I will say "right" or "left" and then I will say "ready." A short time after I say "ready" the light you are to watch will come on.

Are there any questions?

The right and left presentations of $S_1$ were randomly distributed and appeared from .5 to 3 sec. after the ready signal. The intervals at which $S_2$ followed $S_1$ were randomly sequenced. Each $S$ in Group E reacted a total of 380 times, 20 times with $S_2$ presented at each of the 18 intervals and 20 times when $S_2$ was omitted, the latter being randomly interspersed among the trials in which $S_2$ followed the primary signal. The $Ss$ in Group C reacted 360 times to $S_1$ presented in random

order on the right or on the left and with no $S_2$. Two control sets of RTs were thus available with which to compare the experimental RTs: (a) the RTs of Group E without $S_2$, and (b) the RTs of Group C with only $S_1$.

The $Ss$ were given four trials with the primary stimuli to familiarize them with the procedure. There was a 5-sec. break between trials and a 2-min. break at the end of Trial 180. The $Ss$ were not told the purpose of the experiment or the role of $S_2$. All parts of the equipment were shielded from $Ss$ except the light stimuli.

## RESULTS

The results leave no doubt as to the inhibiting effect of $S_2$ on RT to $S_1$. The results of a trend analysis (Grant, 1956) of the data across interstimulus intervals in Table 1 show that the quadratic component is significant at the .05 level and that the linear and cubic components are not significant.[2] This finding lends support to the expectation that the effect of $S_2$ should be minimal at some very short interval following $S_1$, that it should increase to some maximal value or values at certain interstimulus intervals, and then decline as the response is being consummated. Both this reasoning and the trend analysis suggest a parabola as the proper type of function to fit the data in Fig. 1. A second finding from the trend analysis is that the overall influence of interstimulus intervals is significant at the .01 level, which is confirmed by individual $t$ tests of the differences between the mean RTs of Group E with and without $S_2$: 17 of the 18 interstimulus intervals yield significantly longer RTs at the following levels: 9 beyond .001, 4 beyond .01, 4 between .01 and .05, and 1 not significant (with a 180-msec. interval). The $t$ tests were based on comparisons of means of 20

TABLE 1

TREND ANALYSIS OF REACTION TIME × INTERVALS DATA

| Source of Variation | $df$ | $MS$ | $F$ |
|---|---|---|---|
| Linear | 1 | 26,332 | 2.183 |
| Quadratic | 1 | 74,077 | 6.142* |
| Cubic | 1 | 2,533 | .21 |
| Intervals | 17 | 44,653 | 3.703** |
| $Ss$ | 9 | 1,292,452 | |
| $Ss × I$ | 153 | 12,060 | |
| Residual | 14 | 45,169 | 3.745** |

* $P = .05$.
** $P = .01$.
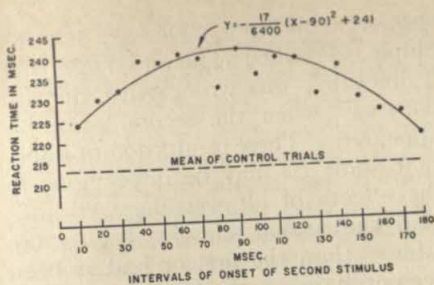
FIG. 1. Reaction time to a primary stimulus when a second stimulus follows at intervals ranging from 10 to 180 msec.

RTs by 10 $S$s under the two experimental conditions using 9 $df$ at each of the 18 interstimulus intervals. A second control, the mean RT of Group C which was given only $S_1$, was even lower than the mean RT in the control trials of Group E when $S_2$ was omitted ($t = 2.82$, $P < .01$), a finding that shows that there was some carry-over in Group E from the $S_2$ trials to the trials in which only $S_1$ was presented.

The curve in Fig. 1 represents a parabola made to fit the 10-, 90-, and 180-msec. intervals since it is at these times that the minimal and maximal inhibiting effects of $S_2$ are found under the conditions of this experiment. The RTs at 80, 100, and 130 msec. depart quite widely from the fitted curve and could undoubtedly have been approximated more closely if it had been fitted according to the least squares criterion which minimizes the total sum of squared differences between all observed and theoretical points. The continuous line in Fig. 1 is given by the equation:

$$Y = (-17/6400)(X - 90)^2 + 241$$

where $Y$ is RT and $X$ is the interval between $S_1$ and $S_2$. Solving this equation for 0 and 180 msec. yields a value of 220 msec. which is higher than the control mean of Group E (213 msec.). However since some $S$s still gave significantly longer RTs

when $S_2$ appeared 180 msec. after $S_1$ and in view of the lower mean RT of Group C as compared with the control RTs of Group E, it is likely that there is still some effect of $S_2$ even as long as 180 msec. after the onset of $S_1$. Taking the curve as a whole we find that the maximal effects are obtained from 40 to 140 msec. following $S_1$ for it is in this region that the parabola is fairly flat. Variations in luminance, hue, or other characteristics of $S_2$ and/or $S_1$ would undoubtedly change the region of maximal effect.

The inhibiting effect of $S_2$ appears even more clearly when RTs of Groups E and C are plotted as a function of trials in Fig. 2. The wide separation between the curves for Group E when $S_2$ is given and Group C for all blocks of trials attests the strength of the inhibiting effect of $S_2$. Indeed, the two curves are more widely separated on the whole during the last 180 trials than during the first 180 trials. Analysis of variance showed these two blocks to be significantly different ($P < .05$). On the other hand, the plot of the trials in
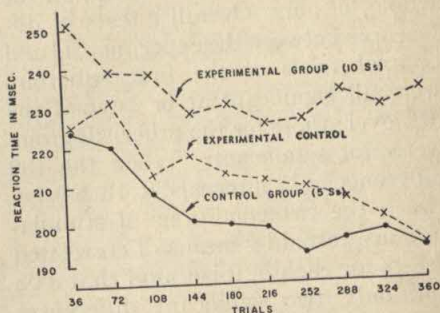


FIG. 2. Decrease in RT in Groups E and C as a result of practice. (Group C improves to a greater extent than does Group E when $S_2$ is present showing that repetition does not counteract the inhibiting effect of the interpolated stimulus, but the RTs of Group E when $S_2$ is omitted as shown by the "Experimental Control" curve, are identical with those of Group C after 360 trials.)

which only $S_1$ was given Group E (intermediate curve in Fig. 2) shows that with practice Ss in Group E were able to react as quickly to a single stimulus as were the Ss in Group C toward the end of the 360 trials. There is, therefore, a differential effect of practice in the single-stimulus condition and in the two-stimulus condition: the carry-over from double stimulation to single-stimulus condition is counteracted by practice, but practice does not overcome the inhibiting effect of $S_2$ in the two-stimulus condition. These findings point to nonattitudinal processes as the basis for the inhibiting effects of $S_2$. Finally, it should be noted that the inhibiting effect of $S_2$ is negated to some extent because of the drop in the Group E curve with the two-stimulus condition from Trial 36 to Trial 144 after which there seems to be no further practice effect.

Individual differences in the extent to which Ss are affected by the interpolated stimulus are striking. Group E divides into two subgroups, one of which is markedly and definitely affected by $S_2$ while the other, on the whole, is not. Overall $t$ tests of differences between the experimental and control trials of the first subgroup were all significant at or beyond the .01 level; those for the other subgroup were not significant. When the 180 differences (18 intervals × 10 Ss) between the two conditions of stimulation are tested by means of correlated $t$ tests for each $S$, it is found that 3 Ss had only two significant differences and 1 $S$ had only three significant differences whereas the other 6 Ss had 11, 13, 14, 14, 16, and 17 significant differences out of a possible 18. It thus appears that some Ss are very much more affected by the interpolated stimulus than are others. Comments made by some of the Ss

support this view. One $S$, for example, remarked again and again that "something was wrong with the apparatus" when the second stimulus appeared. These results do not imply that some Ss may be impervious to the effects of all secondary stimuli: Had the second stimulus been more intense than the first or had it been presented in another sense modality or otherwise made more impressive, it is highly probable (indeed certain, in our minds) that they would have given significantly longer RTs. Even these 4 Ss, as pointed out above, did have significantly longer RTs on some of the interpolated intervals.

## DISCUSSION

The data of this experiment seem to establish very clearly that a stimulus following the primary signal to react has the effect of lengthening RT according to a parabolic function, the effect being minimal 10 msec. after the onset of the primary stimulus, increasing to about 40 msec., and having maximal effect from 40 msec. to 140 msec. after which RT approaches the single-stimulus condition and is not statistically different from RT with one stimulus at the 180-msec. interval. The inhibiting effect of the second stimulus is found even after 360 repetitions although learning has clearly occurred in the single-stimulus condition as shown by the decline in the RT × Trials curve (Fig. 2) for both Group C and Group E. It appears unlikely that higher order, attitudinal, or other perceptual-cognitive processes can be responsible for the effect described here. Rather it seems more probable that lower order, automatically acting mechanisms not under voluntary control, are at work. The difficulty in explaining this phenomenon springs from the fact that a stimulus coming when the response is almost completed can exercise an inhibiting effect. At first sight it might appear that we are dealing with a higher order mechanism in that 4 of the 10 Ss

were not significantly affected by $S_2$ but we saw that even these $S$s had significantly longer RTs on 2 or 3 of the 18 intervals than on their control trials. We confidently expect that such $S$s would be influenced by a second stimulus if it were made more intense than the first or if a more compelling condition were used than was the case in this experiment.

An explanation, in physiological terms, at this time must be purely ad hoc and speculative. Two hypotheses, among the many we have considered, may be ventured as follows: (a) The first hypothesis involves the assumption of *inhibitory* fibers that are *faster* than the ordinary motor fibers innervating voluntary responses. Such fibers might be analogous to slow and fast afferent fibers that are known to function in mediating *slow* and *fast* pain or to differences in speed of afferent conduction associated with differences in fiber size. Whether or not such fast, inhibitory *efferent* fibers exist we do not know and if none have been discovered these data warrant a search for them by electrophysiological means. (b) The second hypothesis, while less suggestive of specific neurological or physiological mechanisms, has behavioral implications and may be stated as follows: Let us assume that the reflex arc is a total, ongoing process such that a disturbance in any part of it disrupts the ongoing activity with the result that a new integration is required for the activity to be resumed or completed. On this basis $S$ is set to react to a single stimulus and when the second stimulus appears, even though he has not been instructed to react or attend to it, nevertheless it breaks into the ongoing response with consequent lengthened RT. It is evident that an hypothesis that explains the inhibiting effect of a second stimulus after a very short interval, e.g., 10–40

msec., may not explain its effect after a comparatively long interval, e.g., 170 msec., or vice versa. An hypothesis must explain lengthened RTs found with all intervals used in this study.

## SUMMARY

An earlier finding that a light following a sound or a sound following a light after an interval of 75 msec. lengthened RT to the stimulus first presented was verified in the present experiment in which the stimuli were both visual. RT to $S_1$ was significantly increased when $S_2$ followed $S_1$ at intervals ranging from 10 to 170 msec. with the maximum effect occurring from 40 to 140 msec. Trend analysis showed the quadratic components in the Intervals $\times$ RT data to be significant. While RT with $S_2$ present decreases with practice, the greater improvement of control $S$s as compared with experimental $S$s shows that the inhibitory effect of $S_2$ on $S_1$ is not completely negated by 360 repetitions. Four of the 10 experimental $S$s failed to give significantly lengthened RTs over all 18 intervals although they did have significantly higher RTs at some of the intervals. While the individual differences and results of practice seem to argue in favor of attitudinal factors as responsible for the effect, other facts argue against this explanation. Two hypotheses were discussed but neither seems completely satisfactory to explain all the facts.

## REFERENCES

GRANT, D. A.   Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

TEICHNER, W. H.   Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

TODD, J. W.   Reaction to multiple stimuli. *Arch. Psychol.*, N. Y., 1912, Whole No. 25.

# RESPONSE SUPPRESSION IN PERCEPTUAL DEFENSE [1]

## ROBERT B. ZAJONC

### *University of Michigan*

Recent theorizing maintains that the phenomenon of perceptual defense can be accounted for in terms of response processes. Elevated thresholds to taboo words are now generally regarded as reflecting a response bias deriving from either previous experience (Goldiamond & Hawkins, 1958), set (Postman, Bronson, & Gropper, 1953), or conflict (Brown, 1961), rather than a defensive perceptual blocking. Although the core of the issue deals with the relative contributions of the stimulus and response to the perceptual defense effect, studies attempting to evaluate such relative contributions have been rather few (Matthews & Wertheimer, 1958; Neisser, 1954). It is the purpose of this experiment to determine the extent to which both recognition threshold and the galvonic skin response (GSR) are influenced by the stimulus and to what extent by the response, using a procedure first suggested by Garner, Hake, and Eriksen (1956). First, threshold and GSR data for a set of taboo and neutral words were obtained by means of standard methods. Secondly, a paired-associate list was constructed using the previously exposed words as stimulus terms and a new set of taboo and neutral words as response terms. Some taboo stimuli were paired with taboo, others with neutral response terms. Neutral stimuli too,

were sometimes paired with neutral and sometimes with taboo response terms. All Ss learned the paired-associate list to a criterion. The third step consisted of a repeated threshold and GSR assessment of the original stimuli. Now, however, one group was required to indicate recognition as before, i.e., by reading out loud the word presented tachistoscopically, and another by saying the appropriate response term which they have learned in the previous paired-associate task. Thus, the second group was given an opportunity to indicate recognition by means of responses whose emotional significance was either positively or negatively correlated with the emotional significance of the stimulus.

## METHOD

*Subjects.*—Forty male Ss, all enrolled at the University of Michigan, participated in the experiment. They were randomly assigned to two experimental groups consisting of 20 Ss each. The Ss were paid $1.25 per hour for participation in the experiment.

*Apparatus.*—Gerbrands' transparent mirror tachistoscope with an instant start fluorescent lamp circuit was employed to present stimuli. Skin resistance changes were observed by means of a Lafayette psychogalvanometer Model 603-A.

*Materials.*—Stimulus words were printed in black 2-in. block letters and presented in the center of the exposure field on gray 10×12 in. cards (54.5% reflectance). Stimulus-response pairs in the paired-associated task were shown in the same manner. Twelve taboo and 12 neutral words were selected from McGinnies' (1949) original list of 18 words to which equivalent neutral and taboo stimuli were added. Half of the taboo and half of the neutral words were used as stimuli in the threshold assessment and as stimulus terms in the paired-associate training task. The remainder of the list was used as response

terms in the paired-associate task. Three taboo stimuli were paired with three taboo responses, three taboo stimuli were paired with three neutral responses, three neutral stimuli were paired with three taboo responses, and three neutral stimuli with three neutral responses. These sets of words will be referred to as the TT, TN, NT, and NN sets. The 24 words were APPLE, BROOM, CANDY, CHAIR, CHILD, FLOOR, MUSIC, RAINS, RIVER, SHELF, STOVE, TRADE, BALLS, BELLY, BLEED, FAIRY, FILTH, HYMEN, KOTEX, PENIS, PUBIC, RAPED, VOMIT, WHORE.

*Procedure.*—As briefly outlined above, the procedure consisted of two recognition threshold and GSR assessment sessions separated by an intervening paired-associate learning task. The Ss were divided into two groups of 20 Ss each, one of which was required during the second threshold assessment session to indicate recognition in terms of the stimuli presented (Group S), the other in terms of the response terms paired with the stimuli (Group R).

Thresholds were obtained by the ascending method of limits in .01-sec. steps beginning with .05 sec. below S's threshold to a neutral training word. The intertrial intervals were approximately 30 sec. Some of the words were shown more than once in order to eliminate prerecognition guesses during the second threshold assessment, when full knowledge of the list was already available to Ss. For those words data from the first presentation alone were included in the analysis. The criterion of threshold was the first correct recognition of the word in Group S and the first emission of the correct response term in Group R.

The S was seated with his head against the eyepieces and with his hand to which electrodes were affixed lying relaxed on the table. A rest period of 1 min. was given after the first threshold assessment session and after the paired-associate training.

The tachistoscope was operated by an adult male and the psychogalvanometer by an adult female. The GSR readings were taken in terms of reduction in resistance from the basal resistance level, which was adjusted for each stimulus exposure. Only those reactions which occurred within 5 sec. following stimulus exposure, and only those for which the resistance returned to the immediate neighborhood of the pre-exposure level were recorded. GSR readings were taken on every presentation of the stimulus word. Since some Ss recognized the word on the fourth exposure, only two prerecognition trials and the recognition trial were

## TABLE 1

MEAN RECOGNITION THRESHOLDS (SEC.) BEFORE PAIRED-ASSOCIATE TRAINING

| Group | Later PA Conditions | | | | Words | |
| | TT | TN | NT | NN | All Taboo | All Neutral |
|---|---|---|---|---|---|---|
| S | .219 | .228 | .193 | .198 | .223 | .196 |
| R | .234 | .233 | .206 | .213 | .234 | .210 |
| Both | .227 | .231 | .200 | .206 | .229 | .203 |

considered. Thus, for each S three GSR scores were computed for each set of stimuli, and for the purposes of analysis all were converted to standard scores with a mean of 50 and SD of 10 for all 40 Ss.

The paired-associate task was conducted using a 2-sec. interval for the presentation of the stimulus and a 2-sec. interval for the presentation of the pair, with 20 sec. between trials. All terms were presented tachistoscopically. The order of the stimuli was randomly altered from trial to trial. Three consecutive correct anticipations of the entire list were used as the criterion.

## RESULTS

*Recognition Threshold and GSR before Paired-Associate Training*

Mean recognition thresholds obtained before paired-associate training are shown in Table 1. The analysis of variance for these results showed that the only significant effect is due to the difference between taboo and neutral words ($F = 17.08$, $P < .001$). Although the mean recognition thresholds for Group R are somewhat higher than those for Group S, this difference is not significant. Also, no significant differences were obtained between taboo words to be later used with taboo responses (TT) and taboo words to be later used with neutral responses (TN). Nor was there any difference between neutral words to be later used with taboo responses (NT) and neutral words to be later used with neutral responses (NN).
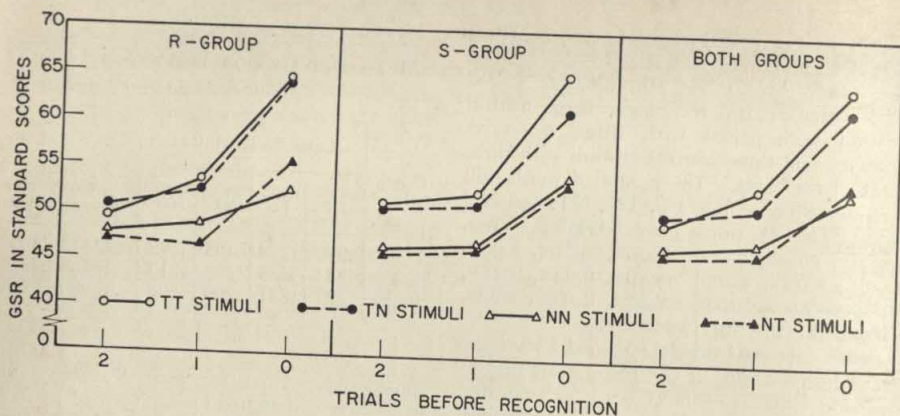
FIG. 1.  GSR before paired-associate training.

The GSRs are shown in Fig. 1. Again no difference between the experimental groups was found. It is evident from the results that on all trials taboo words exceed neutral words in GSR ($F = 30.99, P < .001$). It is also clear that there is a considerable rise in the GSR on the recognition trial ($F = 32.94, P < .001$). No significant differences between TT and TN words as well as between NN and NT words were found for either of the two groups.

*Paired-Associate Learning*

Average number of trials to learn the four sets of associations are presented in Table 2. The means represent the number of trials which $Ss$ required to learn a given association to a criterion of three correct anticipations, averaged for the three items in each set. Shown in Table 2 is also the average number of errors for each set of pairs. The results indicate that the four sets of associations were not learned at the same rate (see Table 3). In particular, the TN pairs seem to be the most difficult, and the TT easiest. The analysis of variance presented in Table 3 shows a significant effect due to differences between word sets,

which is primarily due to the type of response. In general, pairs with a taboo response require fewer trials and lead to fewer errors than pairs with neutral responses. Of particular importance to the present experiment is the difference between the TN and NT pairs. If speed of learning and number of errors reflect the degree to which a given response has become attached to the stimulus word, the TN stimuli should, during the subsequent threshold and GSR assessment of Group R, be more handicapped than NT stimuli. The differences between these pairs in both trials to criterion and average number of errors are significant at the .001 level.

*Recognition Threshold and GSR after Paired-Associate Training*

*Group R.*—Table 4 shows recognition thresholds for the four sets of words for the condition in which $Ss$ indicated recognition by means of the response term acquired during the paired-associate training. It is apparent that, compared with those obtained before the paired-associate training, the thresholds to all the words are considerably reduced. It is

TABLE 2

MEAN TRIALS AND ERRORS TO CRITERION IN PAIRED-ASSOCIATE LEARNING

| Group | Pairs | | | | Words | | | |
|---|---|---|---|---|---|---|---|---|
| | TT | TN | NT | NN | Taboo Stimuli | Neutral Stimuli | Taboo Responses | Neutral Responses |
| Group S Trials Errors | 4.56 4.55 | 5.62 7.10 | 4.37 3.70 | 5.13 6.10 | 5.09 5.83 | 4.75 4.90 | 4.47 4.13 | 5.38 6.60 |
| Group R Trials Errors | 4.21 2.90 | 4.95 5.00 | 4.48 3.80 | 4.28 3.55 | 4.58 3.95 | 4.38 3.68 | 4.35 3.35 | 4.62 4.28 |
| Both Trials Errors | 4.39 3.73 | 5.29 6.05 | 4.43 3.75 | 4.71 4.83 | 4.84 4.89 | 4.57 4.29 | 4.41 3.74 | 5.00 5.44 |

also clear that no longer does the recognition threshold totally depend on the stimulus. There is a considerable effect due to the response which S utilizes in indicating recognition. It should be pointed out that S's ability to give evidence of recognition, not by means of a word which is presented but by means of a response previously learned, depends on the degree to which these responses were fixated. It will be recalled that the

learning of the four types of associations was not uniform. In particular there was a considerable difference between the TN and the NT pairs, in favor of the latter. Moreover, the examination of the results on paired-associate learning disclosed a significant effect due to individual differences. The $F$ ratios evaluating the individual difference effect were 8.79 for trials to criterion, and 11.41 for errors, which for the degrees of free-

TABLE 3

ANALYSIS OF VARIANCE FOR DATA IN TABLE 2

| Source | df | Trials to Criterion | | Errors | |
|---|---|---|---|---|---|
| | | MS | F | MS | F |
| Treatments (A) | 1 | 8.06 | .83 | 96.09 | 1.09 |
| Words | 3 | 6.69 | 6.03*** | 48.54 | 6.23*** |
| S component (B₁) | 1 | 3.11 | 3.27 | 14.40 | 2.68 |
| R component (B₂) | 1 | 13.40 | 8.93** | 115.60 | 11.73** |
| B₁ × B₂ | 1 | 3.57 | 4.12 | 15.62 | 1.92 |
| Treatments × Words | 3 | 1.78 | 1.60 | 13.45 | 1.73 |
| A × B₁ | 1 | .24 | .25 | 4.23 | .79 |
| A × B₂ | 1 | 3.93 | 2.62 | 24.03 | 2.44 |
| A × B₁ × B₂ | 1 | 1.18 | 1.37 | 12.10 | 1.49 |
| Error (b) | 38 | 9.76 | | 88.91 | |
| Error (w) | 114 | 1.11 | | 7.79 | |
| Ss × B₁ | 38 | .95 | | 5.38 | |
| Ss × B₂ | 38 | 1.50 | | 9.85 | |
| Ss × B₁ × B₂ | 38 | .86 | | 8.13 | |

** $P = .01$.
*** $P = .001$.

TABLE 4

MEAN RECOGNITION THRESHOLDS (SEC.) IN GROUP R AFTER PAIRED-ASSOCIATE LEARNING

| Group | PA Pairs | | | | Words | | | |
|---|---|---|---|---|---|---|---|---|
| | TT | TN | NT | NN | Taboo Stimuli | Neutral Stimuli | Taboo Responses | Neutral Responses |
| Rapid learners | .151 | .138 | .145 | .138 | .145 | .142 | .148 | .138 |
| Slow learners | .178 | .180 | .174 | .172 | .179 | .173 | .176 | .176 |
| All Ss | .165 | .159 | .160 | .155 | .162 | .158 | .163 | .157 |
| Adjusted means for all Ss | .168 | .153 | .160 | .157 | .161 | .159 | .164 | .155 |

dom given are significant well beyond the .001 level. We would expect a more reliable test of the relative contributions of the stimulus shown and of the response given from Ss who learned these responses well. Group R was therefore divided at the median number of trials to criterion, and the recognition thresholds for the rapid and slow learners are shown in Table 4, and the analysis of variance in Table 5. It is clear from Table 4 that slow learners manifest considerably higher recognition thresholds for *all* the words. The difference between groups is significant at better

TABLE 5

ANALYSIS OF VARIANCE FOR DATA IN TABLE 4

| Source | df | MS | F |
|---|---|---|---|
| Groups (Rapid vs. Slow) (A) | 1 | 24,945 | 6.72* |
| Words | 3 | 338 | 2.54 |
| S component (B₁) | 1 | 466 | 2.13 |
| R component (B₂) | 1 | 546 | 5.00* |
| B₁ × B₂ | 1 | 1 | <1.00 |
| Groups × Words | 3 | 131 | <1.00 |
| A × B₁ | 1 | 35 | <1.00 |
| A × B₂ | 1 | 536 | 4.91* |
| A × B₁ × B₂ | 1 | 122 | 1.67 |
| Error (b) | 18 | 3,711 | |
| Error (w) | 54 | 133 | |
| Ss × B₁ | 18 | 218 | |
| Ss × B₂ | 18 | 109 | |
| Ss × B₁ × B₂ | 18 | 73 | |

* P = .05.

than the .05 level. It appears that the slow learner's recognition threshold depends primarily on the type of stimulus presented, while that of rapid learners on the response which they were required to make. However, the Groups × Stimulus × Response interaction was not significant. The overall results, however, indicate that the effects due to the stimulus component were not significant while those due to the response were significant. Further support for the conclusion that recognition threshold depends primarily on the type of response required is obtained when the data are adjusted for differences in learning the four types of associations. The mean recognition thresholds, adjusted by means of the regression equation relating the former to the number of trials, are shown at the bottom of Table 4. Analysis of covariance performed on these results disclosed a significant effect due to the response component ($F = 8.78$, $df = 1/17$) and no effects due to stimulus.

The GSR data shown in Fig. 2 follow a similar pattern. Again, as compared with the results obtained before paired-associate training, the GSRs are weaker. The analysis of variance (Table 6) shows a significant effect due to the differences between

FIG. 2. GSR after paired-associate training (Group R).

words which seems to be a function of the stimulus and of the response component as well. The results of both groups combined indicate that on the second prerecognition trial the GSRs do not follow any particular

TABLE 6

ANALYSES OF VARIANCE FOR DATA IN FIG. 2 AND 3

| Source | df | Group R | | Group S | |
|---|---|---|---|---|---|
| | | MS | F | MS | F |
| Groups (Rapid vs. Slow) (A) | 1 | 522.4 | 1.19 | 161.7 | <1.00 |
| Words | 3 | 171.1 | 6.35** | 78.6 | 3.48* |
| S component (B₁) | 1 | 149.6 | 7.03* | 227.0 | 7.73* |
| R component (B₂) | 1 | 356.5 | 10.57** | 2.5 | <1.00 |
| B₁×B₂ | 1 | 7.3 | <1.00 | 6.3 | <1.00 |
| Trials (C) | 1 | 2,235.6 | 28.97*** | 942.5 | 12.57*** |
| Groups×Words | 2 | 9.6 | <1.00 | 26.1 | 1.16 |
| A×B₁ | 3 | .3 | <1.00 | 16.9 | <1.00 |
| A×B₂ | 1 | 27.8 | <1.00 | 51.4 | 1.92 |
| A×B₁×B₂ | 1 | .8 | <1.00 | 10.1 | <1.00 |
| Words×Trials | 1 | 136.7 | 5.44*** | 38.6 | 3.08** |
| B₁×C | 6 | 87.8 | 4.27* | 79.0 | 4.71* |
| B₂×C | 2 | 319.6 | 12.47*** | 6.6 | <1.00 |
| B₁×B₂×C | 2 | 2.8 | <1.00 | 30.4 | 4.94* |
| Groups×Trials | 2 | 25.3 | <1.00 | 39.4 | <1.00 |
| Groups×Words×Trials | 6 | 32.6 | 1.30 | 22.2 | 1.77 |
| A×B₁×C | 2 | 77.5 | 3.77* | 36.7 | 2.19 |
| A×B₂×C | 2 | 5.7 | <1.00 | 4.8 | <1.00 |
| A×B₁×B₂×C | 2 | 14.7 | <1.00 | 25.1 | 4.09* |
| Error (b) | 18 | 439.1 | | 441.1 | |
| Error (w)₁: Ss×Words | 54 | 27.0 | | 22.6 | |
| Ss×B₁ | 18 | 21.3 | | 29.4 | |
| Ss×B₂ | 18 | 33.7 | | 26.8 | |
| Ss×B₁×B₂ | 18 | 25.9 | | 11.6 | |
| Error (w)₂: Ss×Word×Trials | 36 | 77.2 | | 75.0 | |
| Error (w)₃: Ss×Words×Trials | 108 | 25.1 | | 12.6 | |
| Ss×B₁×C | 36 | 20.6 | | 16.8 | |
| Ss×B₂×C | 36 | 25.6 | | 14.7 | |
| Ss×B₁×B₂×C | 36 | 29.2 | | 6.2 | |

* P < .05.
** P < .01.
*** P < .001.

TABLE 7

MEAN RECOGNITION THRESHOLDS (SEC.) IN GROUPS AFTER PAIRED-ASSOCIATE TRAINING

| Group | PA Pairs | | | | Words | | | |
|---|---|---|---|---|---|---|---|---|
| | TT | TN | NT | NN | Taboo Stimuli | Neutral Stimuli | Taboo Responses | Neutral Responses |
| Rapid learners | .153 | .151 | .143 | .141 | .152 | .142 | .148 | .146 |
| Slow learners | .167 | .175 | .154 | .162 | .171 | .158 | .160 | .168 |
| All $Ss$ | .160 | .163 | .149 | .152 | .161 | .150 | .154 | .157 |

pattern. However, the curves for the rapid learners show a pattern of particular interest. On the second prerecognition trial the GSRs seem to depend primarily on the stimulus component; their order is TN, TT, NN, and NT. As the $Ss$ approach recognition the stimulus effect is gradually replaced by the response effect and the GSRs are ordered according to the response. One may interpret this result to mean that stimulation present two trials before recognition is probably too weak to call out strong anticipatory partial responses. As soon as the stimulation gains in strength and becomes capable of evoking some parts of the learned

TABLE 8

ANALYSIS OF VARIANCE FOR DATA IN TABLE 7

| Source | df | MS | F |
|---|---|---|---|
| Groups (Rapid vs. Slow Learners) (A) | 1 | 10,160 | 3.12 |
| Words | 3 | 896 | 3.03 |
| S component ($B_1$) | 1 | 2,532 | 5.34* |
| R component ($B_2$) | 1 | 157 | <1.00 |
| $B_1 \times B_2$ | 1 | 0 | <1.00 |
| Groups × Words | 3 | 150 | <1.00 |
| $A \times B_1$ | 1 | 131 | <1.00 |
| $A \times B_2$ | 1 | 419 | 1.68 |
| $A \times B_1 \times B_2$ | 1 | 1 | <1.00 |
| Error (b) | 18 | 3,088 | |
| Error (w) | 54 | 296 | |
| $Ss \times B_1$ | 18 | 474 | |
| $Ss \times B_2$ | 18 | 250 | |
| $Ss \times B_1 \times B_2$ | 18 | 164 | |

* $P = .05$.

response, the autonomic reactions lose their dependence upon the stimulus and begin to be dominated by the response component.

The mean GSR reaction for rapid learners was 47.55 and for slow learners 50.53, but as is evident from Table 6 this difference was not significant.

*Group S.*—The principal purpose of the paired-associate learning task was to enable $Ss$ to give evidence of recognition of the stimulus words without having to say them. However, it is possible to argue that the training simultaneously produced temporary changes in the emotional quality of the stimulus words. Thus, taboo stimuli which were paired with neutral responses could, by virtue of the repeatedly reinforced association, have become emotionally "neutralized." Similarly, conditioning a taboo response to a neutral stimulus word might have affected the emotional quality of the latter. These eventualities are of course quite remote because of the small number of conditioning trials involved. If conditioning of the type suggested has in fact taken place then the recognition thresholds and the GSR data should show the same patterns in Groups S and R. The average recognition thresholds for Group S are shown in Table 7, and the analysis of variance for these results has shown

effects only due to the stimulus component. It is of interest to note that as was the case in Group R slow learners in Group S also showed somewhat higher recognition thresholds than rapid learners. However, this difference failed to reach an acceptable level of significance.

Neither do the GSR results shown in Fig. 3 suggest any conditioning effect. Besides the increase in reactions over trials, the only significant effect is that due to the stimulus component. The analysis of variance in Table 8 shows an $F$ ratio significant at the .05 level for the stimulus component. On the trials immediately preceding recognition there is a slight but not significant response effect for rapid learners. Also, as observed before, the GSRs of rapid learners are somewhat less than those of slow learners (45.69 and 47.16, respectively), but this difference is decidedly not reliable.

## DISCUSSION

The evidence presented failed to disclose perceptual effects of any significance. The recognition threshold was found to be a function not of what $S$ saw but what he had to say. Moreover, GSR data follow an identical pattern.

The GSRs were found to be produced not by the stimulus alone, but depended primarily on the response required of $S$. The results are best accounted for by Brown's (1961) competing response theory. Irrespective of the stimulus, if the responses were in conflict with an inhibitory tendency, that is, if $S$ had to make a vulgar response, both recognition threshold and GSR were elevated. Stimuli arousing no response conflict failed to produce differential thresholds and GSRs irrespective of their "emotionality." Further support for the response competition hypothesis is seen in the GSR data. In general, the differences in the GSRs were found to increase over trials, reaching their peak upon recognition. To the extent that the GSRs reflect response conflict, one would expect that with increasing exposure time both the positive and the negative tendencies increase, thus generating a stronger conflict.

There is evidence in the data that recognition threshold and GSR are also subject to variation as a result of not only a conflict between a positive and negative tendency, but also as a result of a conflict between competing excitatory tendencies. First we note that both are markedly reduced after familiarization with the stimuli. Before paired-associate learning the response alternatives available to $Ss$ are many, and all of these are in competition. The training
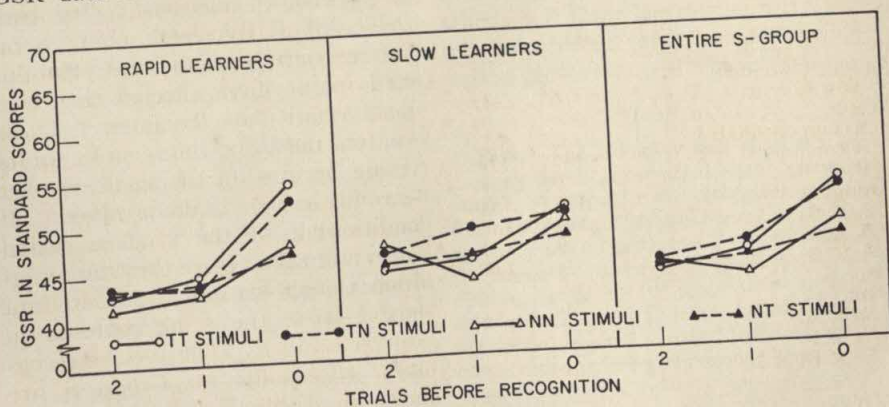


FIG. 3. GSR after paired-associate training (Group S).

reduces them to 12, thus reducing the extent of response competition involved. Secondly, consistent differences in the overall recognition threshold and GSR reactions between the rapid and slow learner were found. If one views the speed of the paired-associate learning and the mean number of errors as an index of the amount of response competition present, these results become quite meaningful.

It is not claimed here that the perceptual defense phenomenon has been disproven. But if the phenomenon is empirically demonstrable its proof must be established by experimental methods other than those commonly used. Perhaps Blum's (1954) forced-choice technique of threshold assessment holds best promise since it eliminates possible effects due to the response process.

## Summary

The role of stimuli and responses in perceptual defense was examined by first obtaining recognition thresholds and GSRs to taboo and neutral words. Subsequently, Ss learned a paired-associate list with the original words serving as stimulus terms and a new set of words as response terms. Half of the neutral stimuli were paired with neutral and half with taboo responses. The same was true of taboo stimuli. Following training, recognition thresholds and GSRs were again measured with one group required to indicate recognition by means of response terms and

another by means of stimulus terms. Both recognition threshold and GSR were found to depend primarily on the response required of the Ss in indicating recognition.

## References

BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **13**, 94–99.

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

MATHEWS, A., & WERTHEIMER, M. A "pure" measure of perceptual defense uncontaminated by response suppression. *J. abnorm. soc. Psychol.*, 1958, **57**, 373–376.

McGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399–402.

POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.

# FACTORS IN THE RETENTION AND RELEARNING OF PERCEPTUAL-MOTOR SKILL [1]

EDWIN A. FLEISHMAN

*Yale University*

AND JAMES F. PARKER, JR.

*Psychological Research Associates, Arlington, Virginia*

In several previous reports, Parker and Fleishman have described studies of complex tracking performance. The first of these (Parker & Fleishman, 1959, 1960) attempted to predict performance at different stages of learning a complex tracking skill. Special efforts were made to predict high levels of proficiency after extensive practice (17 sessions distributed over 6 weeks) with this task. A second study (Parker & Fleishman, 1961) made use of information about the components of tracking skill to facilitate the learning of this skill. The present study is an investigation of factors in the retention and relearning of this same skill.

Previous studies of motor skill retention (e.g., Ammons, Farr, Bloch, Neumann, Dey, Marion, & Ammons, 1958; Battig, Nagel, Voss, & Brogden, 1957; Bell, 1950; Jahnke, 1958; Jones & Bilodeau, 1953; Leavitt & Schlosberg, 1944; Mengelkoch, Adams, & Gainer, 1958; Reynolds & Bilodeau, 1952) all present evidence that continuous control, perceptual-motor skills are well retained over fairly long periods of no practice. What loss occurs appears to be quickly regained. The present study is a more comprehensive study of factors in retention and relearning, using a highly complex continuous control task requiring considerable practice for initial learning.

While essentially a laboratory study, the task was designed to simulate a complex skill, i.e., that of a pilot flying a radar intercept mission. The problem of retention over extended periods without practice is especially critical here. Furthermore, where the skills are of such complexity, the problem of finding the optimum conditions for retraining becomes even more critical.

Specifically, the following questions were investigated. How well is such a skill retained without practice? What is the relation between the length of the "no practice" interval and level of retention? If there is a loss in proficiency, how much practice is required to regain proficiency? What is the relation between retention and level of proficiency after the original learning? Is the *type* of initial training related to retention? What is the relative effectiveness of a distributed vs. massed retraining schedule? Does the type of retraining schedule affect later performance as well as performance during retraining?

## METHOD

*Task.*—The criterion task consisted of a tracking device constructed so as to simulate roughly the display characteristics and control requirements of an air-borne radar intercept mission. The task of S was to maintain the target dot at the center of the oscillograph display, while at the same time nulling a sideslip indicator. That is, S envisioned himself to be flying the attack phase of an airborne radar intercept mission. Thus, if the target was to the right, S made appropriate control movements to steer the craft to the right. These movements would bring him

on target and the dot would return to the center of the display. All turning movements required coordinated action of stick and rudder controls.

Three identical tracking devices were constructed especially for purposes of this study. Photographs and complete schematics of all components are presented elsewhere (Parker & Fleishman, 1959). These devices and related scoring consoles allowed for the testing of from 1 to 3 Ss simultaneously under the control of a single test administrator.

The S's instrument panel contained two displays. The first consisted of a target dot presented on a cathode ray oscillograph. The target course was generated by setting the equation of a swinging pendulum into an analog computer. This produced a sine wave with a frequency of approximately 6 cycles per min. in the horizontal coordinate as the target course. The rate of decay in amplitude was approximately 5% per cycle. However, the dynamic characteristics of the overall task were such that Ss neither perceived this drop in amplitude nor the fact that the target was programed in one coordinate only. Any control imbalances resulted in dot excursions of considerably larger magnitude than those provided by target programing.

Beneath the oscillograph was an inverted 3-in., zero-center voltmeter termed a "sideslip indicator." This meter reading indicated to S a "lack of coordination" in control actions when centering the target dot. This indicator did not constitute an independent task but rather provided S with additional information related to the primary task.

In performing this task Ss used a standard aircraft control system involving a control stick and rudder pedals. These controls were coupled in a manner similar to those of an actual aircraft. Thus, application of right control stick pressure without proper amount of right rudder produced a sideslip to the left and a consequent left deflection on the sideslip indicator.

Control of the target dot in elevation was accomplished by forward and backward movements of the control stick. This was a pure second-order system resulting from the use of two cascaded integrators in the linkage between the control and display. Thus the acceleration of the target dot was directly proportional to stick displacement.

Control of the target dot in azimuth was accomplished by right and left movements of the control stick. This comprised a system involving acceleration control plus an exponential lag network. This dimension was mechanized, using three cascaded integrators with a negative feedback loop around one.

The time constant of this lag network is 1 sec., i.e., it requires 1 sec. to achieve approximately two-thirds $(1-1/e)$ of the final signal resulting from a given stick displacement.

Control of the target dot in azimuth (envisioned as turning the aircraft) and centering the sideslip indicator (coordination display) were both affected by rudder pedal displacement as well as by the control stick. This rudder control of the sideslip indicator involves a simple lag network. Thus, the sideslip indicator displacement was directly proportional to rudder pedal displacement. Movement of the sideslip indicator by stick action represents a velocity control operating through two exponential lag networks.

Movement of the target dot by rudder pedal displacement approximates a pure velocity control.

*Scoring.*—The primary score was the *integrated absolute error score.* This was recorded at the conclusion of every trial and was produced by summing algebraically the three absolute error part scores in accordance with this relationship: $T = 1/2X + 1/2Y + Z$, where $T$ = integrated absolute error score, $X$ = absolute azimuth error, $Y$ = absolute elevation error, and $Z$ = sideslip (lack of coordination) error.

*Initial training.*—The initial learning data upon which this study is based were gathered during the course of two previous studies.[2] In each of these studies Ss spent a total of 17 sessions, distributed over 6 weeks, mastering the tracking task, with each session consisting of 21 1-min. trials.

In the first study (Parker & Fleishman, 1959, 1960) no formal training instructions were administered. Although provided knowledge of results, Ss learned the tracking task "on their own" with the single exception that any questions were answered. This group of Ss will be referred to as Group I.

In the second study (Parker & Fleishman, 1961) 60 Ss, referred to as Group II, spent an identical period of time mastering the same tracking task. These Ss were administered a "common sense" training program. This consisted of an initial explanation and demonstration of the tracking device. This was followed by three sessions of practice with an E monitoring the entire operation and assisting S as required. The remaining 14 sessions consisted of individual practice

[2] The first study was concerned with correlates of performance at different stages of learning this task. The second study evaluated an experimental training program. Group II, in the present study, was the control group in that experimental training program.

followed by critiques with each *S* after Sessions 7, 11, and 15. As would be expected, terminal proficiency for the group trained under this program was significantly superior to that of the group which had no formal training.

*Retention testing.*—Seven groups of 10 *S*s were brought back for retraining following various intervals of no practice. Intervals since training for Group II were 1, 5, 9, and 14 mo. Intervals since training for Group I were 9, 14, and 24 mo. The fact that it was possible to study the 9- and 14-mo. intervals for both groups allows a comparison of retention for the same intervals for two types of original training.

Each retention group was split into two subgroups of 5 *S*s each. One subgroup was retrained during four intensive, continuous retraining sessions[3] during the same day. The other subgroup was retrained during four sessions, each scheduled 1 day apart. The purpose of this experimental breakdown was to allow an evaluation of the relative effectiveness of these two retraining schedules where one involved massed and the other distributed practice.

One week following the final retraining session all *S*s were again tested for one additional session. The purpose of this additional testing was to allow a more adequate evaluation of the two types of retraining. This fifth retraining session was included to show whether any differences which might be found occurred only during the retraining program or whether these differences persisted in later performance. If transfer to later performance could be demonstrated, this could be attributed to differential *learning* during the course of retraining rather than to temporary performance factors (e.g., fatigue, inhibition) during the retraining.

*Matching of retention groups.*—It will be recalled that the seven retention groups were drawn from two groups of original trainees. One group (Group I) had been trained without benefit of specific guidance while the other group (Group II) had been trained with supplementary instruction and guidance. Accordingly, it was found (Parker & Fleishman, 1961) that Group II was superior in terminal proficiency although the number of practice sessions was identical for each group. For the present retention study an attempt was made to match the different retention samples on the basis of final performance level during original learning. This was done separately for the three retention samples drawn from Group I and for the

[3] As before, a "session" includes 21 1-min. trials.

## TABLE 1

FINAL PERFORMANCE LEVELS OF THE SEVEN RETENTION SAMPLES AFTER INITIAL LEARNING

| Retention Interval (Months) | Group I No Formal Guidance | | | Group II Formal Guidance | | |
|---|---|---|---|---|---|---|
| | N | Mean[a] | SD | N | Mean[a] | SD |
| 1 | | | | 10 | 246 | 82 |
| 5 | | | | 10 | 232 | 57 |
| 9 | 9 | 294 | 119 | 9 | 232 | 59 |
| 14 | 7 | 263 | 88 | 9 | 230 | 71 |
| 24 | 8 | 260 | 117 | | | |

[a] Integrated absolute error score.

four retention samples drawn from Group II.

Scores (integrated error) attained by each *S* during the final session of original training were used as a basis for matching. These were converted to standard scores (stanines). An attempt was made to assign a proportionate number from each stanine level to each retention group in order to obtain a normal distribution representative of the original learning population. It soon became apparent that certain *S*s needed to fulfill these requirements were not available for retesting. However, this procedure was followed as closely as possible. A preliminary analysis of variance for the first four samples tested indicated that they were not homogeneous. An adjustment was made by eliminating a few *S*s whose final scores during the original training were extremely poor. This left a total of 62 *S*s in the seven retention samples. Table 1 indicates the number of *S*s in each retention group and the means and *SD*s of their final scores after initial learning. It can be seen that within each of the original learning groups adequate matching was achieved. Analyses of variance performed for each original training group confirmed that the retention samples could be considered comparable (Group I: $F = .22$, $df = 2/21$; Group II: $F = .09$, $df = 3/34$).

The comparability of these retention groups becomes especially apparent when one considers the range of *possible* scores from early to late learning. The curves in Fig. 1 and 2, for example, illustrate this.

## RESULTS

*Magnitude of retention.*—A primary concern is the extent to which the developed performance capability deteriorates through time. However,
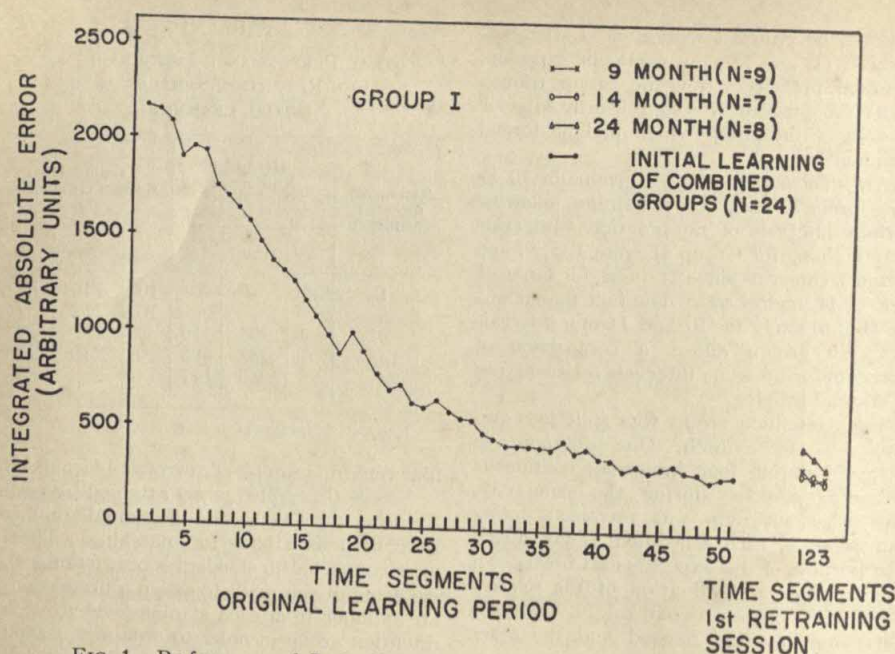
FIG. 1. Performance of Group I (no formal training) during original learning and following varying periods without practice.

a simple comparison of an *S*'s score following some period of no practice with his final score in initial training will not provide a complete understanding of performance loss. One must have information concerning the course of initial learning and the extensiveness of the training required to develop the skill. Figures 1 and 2 present, for Groups I and II, curves illustrating the course of initial learning followed by the results of the first session retention measurements. Each point in the initial learning curve is based on the average of the combined *S*s of the retention samples. Points on the abscissa are directly comparable for original training and retention and represent 6-min. periods within a practice session. Each session consisted of 21 min. of practice. The first three 1-min. trials were not scored to remove warm-up effects.

One of the primary conclusions drawn from Fig. 1 and 2 is that by comparison with the original learning of this skill there is little decrement in performance even for no practice periods of up to 24 mo. There is obviously somewhat more decrement in the 24-mo. group but recovery is rapid even during this first 21-min. retraining session.

*Retention and length of interval.*—In order to obtain a more precise description of performance at the beginning of retraining, the results of the first retraining session were plotted on a trial-by-trial (minute-by-minute) basis. Figure 3 shows that for Group I (no formal training) the major part of performance capability was regained during the first 2 or 3 min. of retraining. It can also be seen that the 24-mo. group is consistently poorer during this first retraining session than are the other two groups. However, even this group is improving to the level of the other groups within this first 21-min. retraining session. A much smaller loss occurs
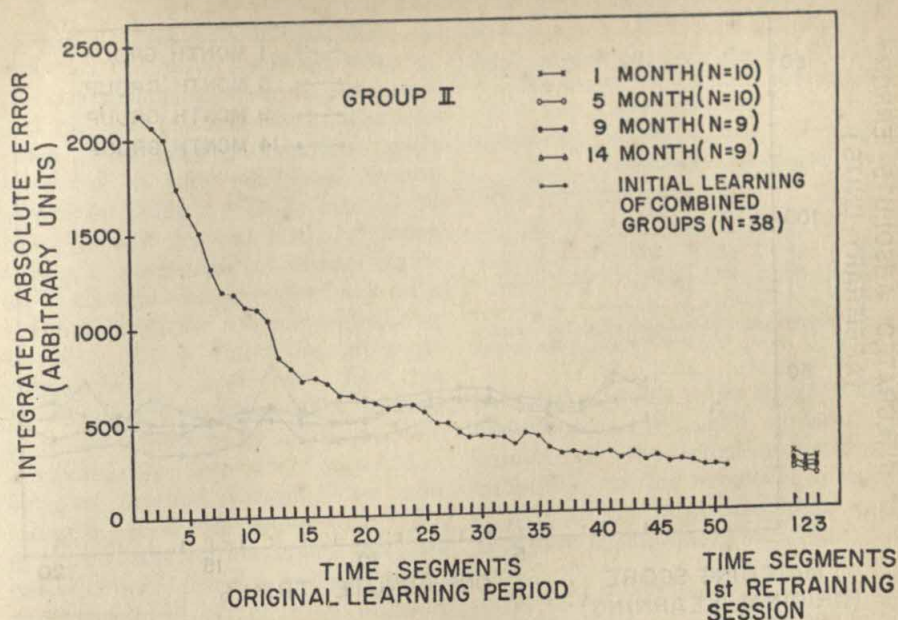
FIG. 2.  Performance of Group II (formal guidance) during original learning
and following varying periods without practice.

for the 9- and 14-mo. intervals and
this is regained in just a few minutes.

Differences between these latter two
groups are negligible.
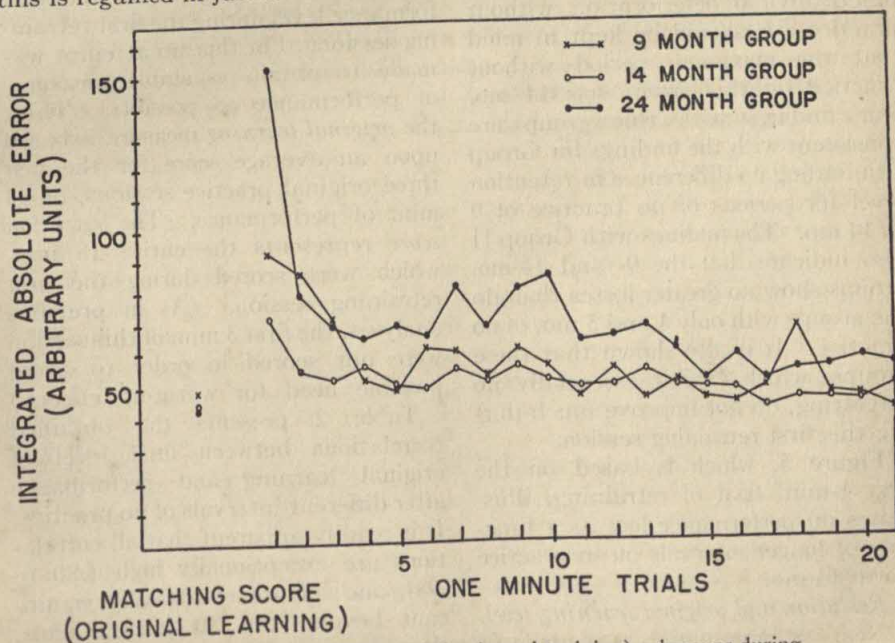


FIG. 3.  Performance of Group I (no formal training) subgroups during
the first retraining session.
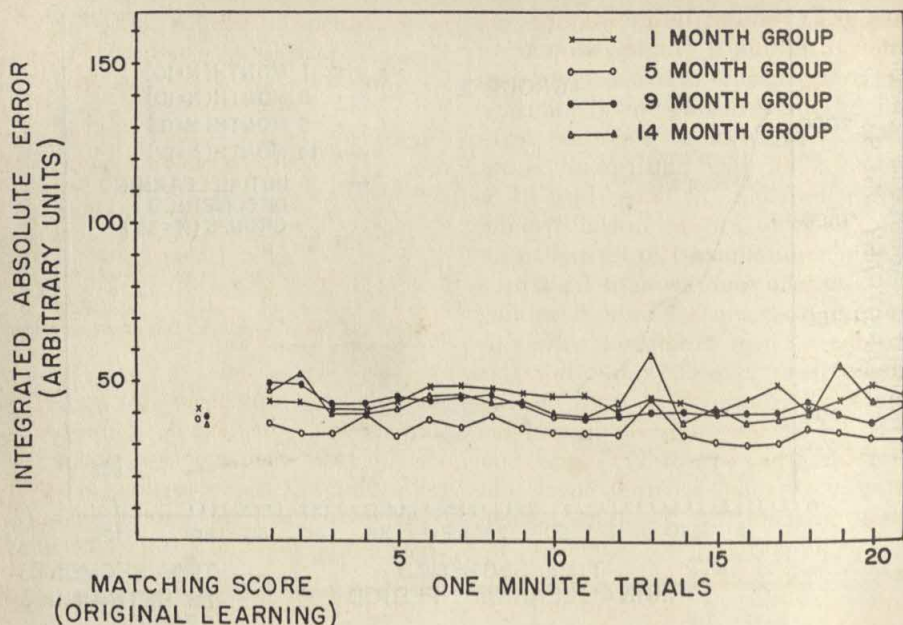
FIG. 4. Performance of Group II (formal guidance) subgroups during the first retraining session

Figure 4 indicates that Group II (formal guidance procedures) showed practically no deterioration without practice. It should be kept in mind that the maximum period without practice for this group was 14 mo. Our findings with this group are consistent with the findings for Group I in finding no differences in retention level for periods of no practice of 9 to 14 mo. The findings with Group II also indicate that the 9- and 14-mo. groups show no greater losses than do the groups with only 1 and 5 mo. of no practice. It is also shown that these groups, which exhibit essentially no forgetting, do not improve much during this first retraining session.

Figure 5, which is based on the first 1-min. trial of retraining, illustrates no performance loss as a function of longer intervals of no practice up to 14 mo.

*Retention and original learning level.* —Next an examination was made of the correlation between final per-

formance level at the conclusion of the original learning period and performance level during the first retraining session. For this, an attempt was made to obtain as stable measures of performance as possible. Thus, the *original learning* measure is based upon an average score for the last three original practice sessions, or 44 min. of performance. The *retention score* represents the entire 18 min. which were scored during the first retraining session. (As in previous analyses, the first 3 min. of this session were not scored in order to avoid possible need for warm-up effects.)

Table 2 presents the obtained correlations between final level of original learning and performance after different intervals of no practice. It is readily apparent that all correlations are exceptionally high (.80 to .98), and all are statistically significant beyond the .01 level. Thus, there is virtually no change in the ordering of *S*s in any group with the

passage of time without practice. In order to obtain a single estimate of the relationship between retention and original learning level, all these cases were pooled together with scores of 40 other additional $Ss$ not shown in Table 2. These last 40 $Ss$ were $Ss$ who learned initially under Group I procedures (no formal training) and who were brought back on a random basis for a single retraining session. Their retention intervals ranged from 9 to 25 mo. For this combined group ($N = 109$), having as it did a wide range of no practice intervals, the zero-order correlation between original learning level and retention score was found to be .80. A partial correlation coefficient between original learning and retention, with the effect of retention interval held constant, was .79. The loss of one point can be attributed to rounding error in the computational process.

The zero-order correlation between retention interval and retention score, for this combined sample of 109 $Ss$ was .30; when initial learning is

partialed out this drops to .23. This underscores the small amount of variance in the retention score attributable to the retention interval relative to the large amount of variance in retention due to initial learning level.

An important question is whether the effect of initial learning upon retention performance is more important following short periods of no practice as opposed to longer intervals. In other words does the relation between initial learning level and retention level dissipate through time?

TABLE 2

CORRELATIONS BETWEEN ORIGINAL LEARNING LEVEL AND RETENTION TEST PERFORMANCE

| Group | Retention Interval (Months) | | | | |
|---|---|---|---|---|---|
| | 1 | 5 | 9 | 14 | 24 |
| I | — | — | .84 | .93 | .80 |
| II | .89 | .85 | .98 | .93 | — |

Note.—All $r$'s are based on $N$s of 10, except for Group I, 9-mo. entry, where $N$ is 9.
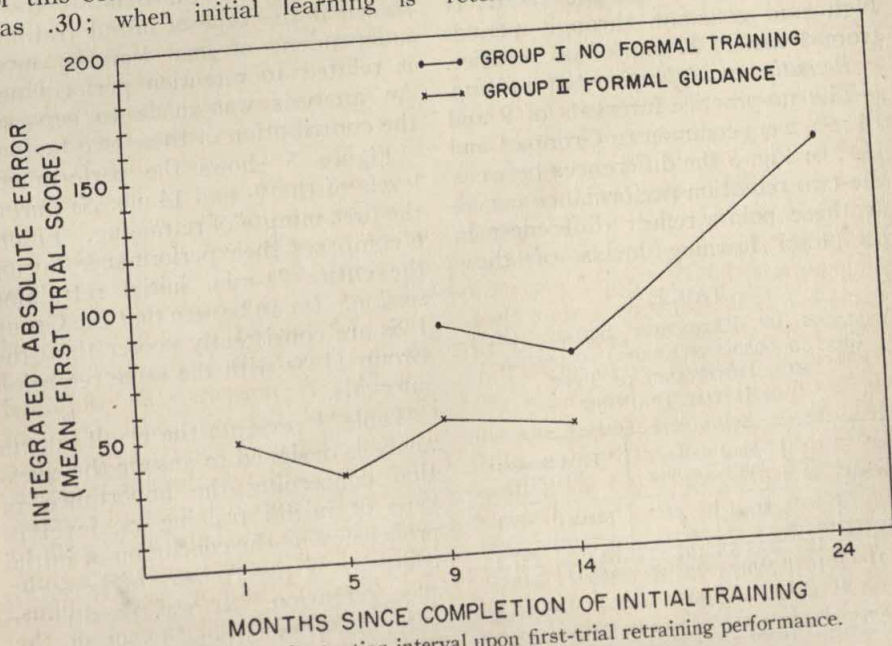


FIG. 5. The effect of retention interval upon first-trial retraining performance.
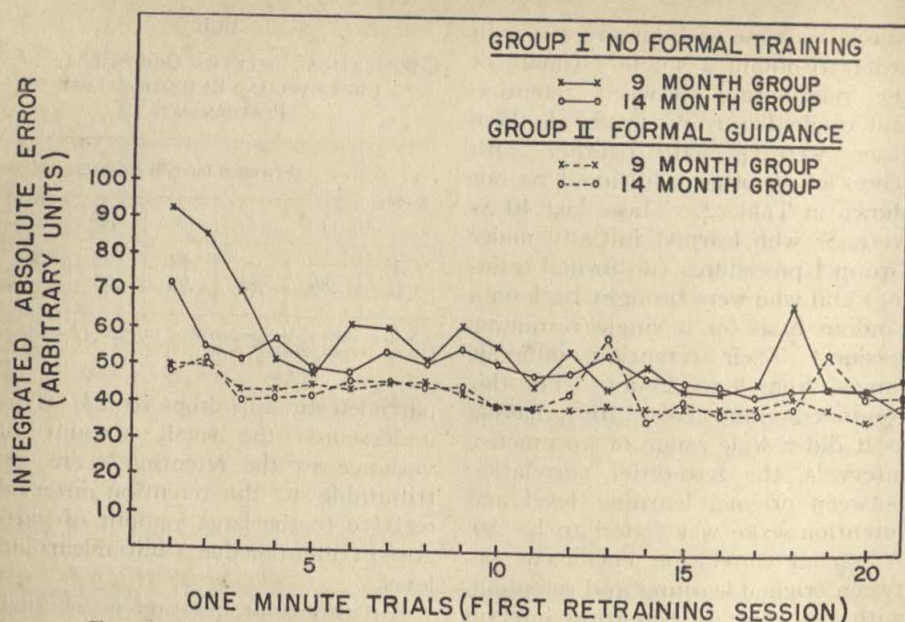
FIG. 6. Comparison of groups with same retention interval, illustrating the importance of prior proficiency level.

The correlations presented in Table 2 offer no evidence in support of this. The relationship between original learning level and retention performance appears to remain relatively high and constant through periods from 1 mo. to 24 mo. of no practice.

*Retention and type of initial training.*—The no-practice intervals of 9 and 14 mo. were common to Groups I and II. In Fig. 5 the differences between the two retention performance curves at these points reflect differences in the final learning levels of these groups. As described earlier, the differences in final learning level result from two types of initial training procedures. It still remains to be shown if the type of initial training, independent of final learning level, is related to retention performance. An analysis was made to separate the contribution of these two factors.

Figure 5 shows the performance levels of the 9- and 14-mo. $S$s during the first minute of retraining. Figure 6 compares their performance during the entire 21-min. initial retraining session. It can be seen that the Group I $S$s are consistently poorer than the Group II $S$s with the same retention intervals.

Table 3 presents the results of an analysis designed to answer the question concerning the importance of type of initial training vs. level of proficiency at the conclusion of initial training, as determiners of performance retention. It was possible to select $S$s from Groups I and II who were matched in terms of retention

TABLE 3

ANALYSIS OF RETENTION SCORES (INTEGRATED ABSOLUTE ERROR) TO DETERMINE IMPORTANCE OF TYPE OF INITIAL TRAINING

| Group | N | Final Session: Initial Training | | First Session: Retention | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| I | 10 | 743.3 | 198.11 | 857.2 | 237.13 |
| II | 10 | 749.2 | 206.20 | 825.9 | 248.64 |

Note.—Subjects matched on retention interval and proficiency at conclusion of initial training.

interval (9 or 14 mo.) as well as in terminal proficiency at the conclusion of initial training. Table 3 presents the matching scores (terminal proficiency after initial training) and the mean scores obtained during the first retention session. No significant difference was found between scores ($t = .32$, $df = 9$). This indicates that the differences among our Groups I and II retention samples following periods of no practice of 9 and 14 mo. are a function of *level of proficiency* at the end of initial training rather than the *type* of initial training used in this study.

*Comparison of retraining[4] procedures.*—Each retention group was split into two subgroups of 5 Ss each.[5] Assignment to a particular group was

[4] To be more accurate, technically, the term "refresher practice" might be used here since no formal training was involved in these practice sessions. For ease of discussion, however, "retraining" is used.

[5] Since certain cases later were dropped in order to obtain matched retention samples, the subgroups which were used in the data analyses ranged in size from 3 to 5 Ss.

on the basis of stanine score at the completion of initial training. This was used as a means of making the subgroups approximately comparable in tracking ability. One group was retrained during four 21-min. sessions with 10 min. rest between sessions; the other during four 21-min. sessions scheduled 1 day apart. Figure 7 presents average performance curves for the retention groups retrained under these two conditions. The varying periods of no practice are equated for the two curves. The two retraining procedures do not result in substantially different performances through the third retraining session, but in the fourth retraining session the distributed practice group appears decidedly superior to the massed practice group.

Apparently, there may be some critical period beyond which performance under massed practice begins to deteriorate. As a means of further evaluating this, the fourth (terminal) session scores for the two retraining procedures were compared statis-
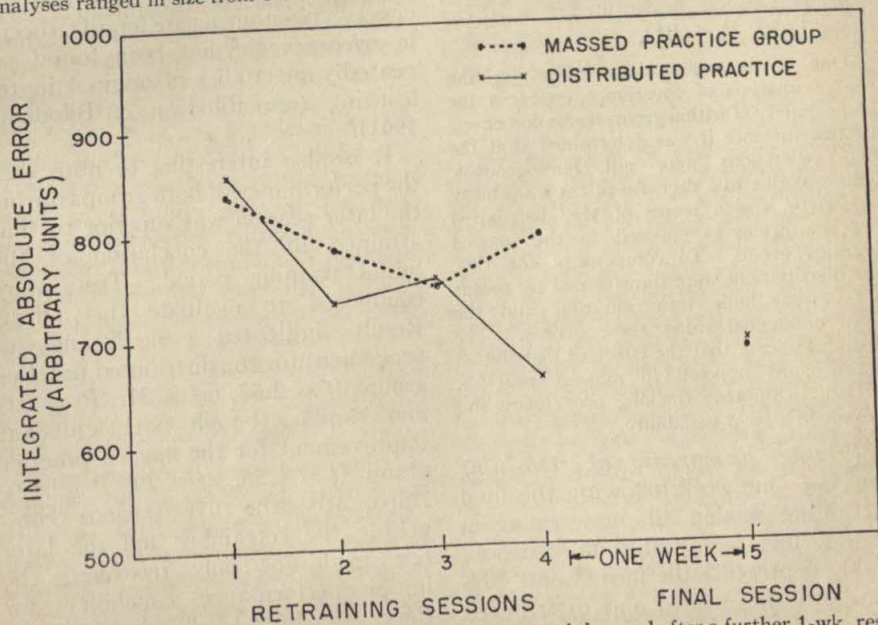


FIG. 7. Effect of different retraining programs during retraining and after a further 1-wk. rest.

TABLE 4

GROUP MEANS AND SDs OF INTEGRATED ABSOLUTE ERROR SCORES DURING FINAL TRAINING SESSION, RETRAINING SESSIONS, AND FINAL RETRAINING SESSION

| Session | Massed Practice ($N = 30$) | | Distributed Practice ($N = 32$) | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Last initial training | 733.4 | 241.9 | 773.4 | 229.0 |
| First retraining | 836.0 | 360.6 | 858.5 | 308.6 |
| Fourth retraining | 801.4 | 365.4 | 659.2 | 232.8 |
| Final (1 wk. later) | 684.5 | 251.0 | 693.8 | 225.0 |

tically. Due to the limited number of cases in each subgroup and the consequent difficulty of matching such scores, an analysis of covariance was conducted which compared fourth session scores while removing the effect of first session scores as a source of variance. In effect, this procedure equates the subgroups statistically and increases the efficiency of the comparison procedure. The results verify the superiority of distributed practice over massed practice as a retraining procedure at the end of four retraining sessions ($F = 10.75$, $df = 1/59$, $P < .01$).

One of the assumptions underlying the use of analysis of covariance concerns the homogeneity of within group regression effect. In this instance it was determined that the two regressions were not homogeneous. This was due to a decrease in the variability of fourth session scores of the distributed practice group as opposed to the massed practice group. To compensate for this, the distributions were transformed by means of a logarithmic transformation and the tests conducted using these scores. The results indicate that the effect of distributed practice, as opposed to massed practice, is both to improve tracking proficiency and to reduce inter-$S$ variability.

*Relative permanence of retraining benefit.*—One week following the final retraining session all $Ss$ were again tested for one additional session. Table 4 presents the means and $SDs$ for both the massed and distributed groups for the first and the final

retraining sessions as well as for the session held 1 wk. later. Whereas group performances initially are approximately equal, at the conclusion of the retraining period the distributed practice group is considerably more proficient. However, 1 week later the two groups again are performing at an approximately equal level and this level is closer to that attained by the distributed practice group at the end of retraining than it is to that of the massed practice group. These differences were evaluated statistically; again, an analysis of covariance procedure was used comparing the two groups for the final session with a control on individual variation during the first retraining session. An $F$ of .005 ($df = 1/59$) clearly indicated that during the final session there was no significant difference between the groups. It appears that the differences observed during retraining are not due to differential *learning*, but to temporary factors affecting performance. Thus the same "massed practice, postrest recovery" phenomena are found to occur in *relearning* as has been found repeatedly in studies of original motor learning (see Bilodeau & Bilodeau, 1961).

It is also interesting to note that the performance of both groups during the later session was superior to that attained at the conclusion of the initial training period. Tests were conducted to evaluate this effect. Results indicated a significant improvement for the distributed practice group ($t = 2.57$, $df = 31$, $P = .02$) and similar though not significant improvement for the massed practice group ($t = 1.56$, $df = 29$, $P = .12$). Apparently the five sessions comprising the retraining and the later test trial not only recovered the initial performance capability for these $Ss$, but produced additional improvement.

*Predicting retention from ability measures.*—Those Ss in Group I had been administered a battery of 44 printed and psychomotor aptitude tests in an earlier study (Parker & Fleishman, 1959, 1960). A subsequent factor analysis of the correlations among these tests identified 15 ability factors, but only 2 of these (Spatial Orientation and Multilimb Coordination) were found related to performance on the tracking task during initial learning. And these two factors, jointly, never contributed more than 25% of the variance in performance at any stage of practice with this task. Nevertheless, it was thought useful to see if measures of these factors were related to performance after periods of *no* practice.

From their loadings on the two factors (see Parker & Fleishman, 1960), the Stick and Rudder Orientation (printed) Test and the Rudder Control (apparatus) Test were chosen to represent the Spatial Orientation and Multilimb Coordination factors, respectively. Correlations between these tests and performance during the first retention session were computed based on an $N$ of 69 (the Group I Ss represented in Table 3, plus the 40 Ss brought back for a single session of retention testing). These zero-order correlations with retention performance were .21 for the Spatial test and .18 for the Coordination test; these coefficients are significant at the .10 but not the .05 level of confidence. To hold the effects of initial learning level and retention interval constant, second-order partial correlations were computed. With these factors partialed out the Spatial test correlated .21 and the Multilimb Coordination test correlated .20 with performance in the retention session. Again these coefficients are significant only at the .10 level, not at the .05 level, for second-order partials.

Thus, for this particular skill, a negligible to insignificant portion of retention performance is attributable to Ss' abilities as measured prior to initial learning.[6] This is true when retention

[6] The distinction between the constructs "ability" and "skill" has been elaborated elsewhere (Fleishman, 1959, 1962; Gagné &

is defined in terms of performance after no practice, as well as when this performance is residualized with respect to initial learning level and retention interval.

Performance on this task during early stages of *initial* learning was shown to be uncorrelated with performance during late stages of initial learning (e.g., as late as Trial 8 the correlation with Trial 50 was only .13); however, practice sessions late in original learning correlated .70 with each other (Parker & Fleishman, 1960). The communality of the final initial learning trial attributable to independently measured ability factors was only .24. Taken together, these findings suggested that proficiency at the end of training was mainly a function of specific habits and skills acquired during the 6 wk. of practice with the task and only to a small extent a function of Ss' abilities prior to his experience with this task.

The present findings indicate this is also true of *retention* performance after prolonged periods of no practice. This is especially apparent when we recall the high correlations (in the .80s and .90s) between proficiency at the conclusion of training and retention performance, relative to the negligible correlations of retention with the independent ability measures.

## SUMMARY

Two groups of Ss were given extended training on a highly complex tracking task. Practice extended over 17 sessions distributed over 6 weeks. The two groups differed only in the amount of verbal guidance provided in initial training. Within each group, subgroups of Ss matched for final proficiency were retested following various no-practice intervals of up to 24 mo. These retention samples were further divided into two subgroups, each of which were given four additional retraining sessions; in one group this relearning practice was massed in 1 day and

Fleishman, 1959). Briefly, "ability" refers to a more general, stable trait of the individual inferred from response consistencies on a given range of tasks. Skill refers to proficiency on a specific task. Some portion of the variance in a given skill can be accounted for in terms of particular component abilities.

for the other group it was distributed over 4 days. One week following the retraining all Ss were retested as a means of evaluating the persistence of the effects of these two relearning schedules.

1. The retention of proficiency in a complex, continuous control, perceptual-motor skill is extremely high, even for no-practice intervals up to 24 mo. For Ss trained initially to high levels of proficiency (Group II), virtually no loss was observed for periods up to 14 mo. What small losses did occur were recovered in the first few minutes of relearning. With 24 mo. of no practice, rapid recovery still occurred during the first 20 min. of relearning.

2. Variations in retention interval from 1 to 14 mo. are shown to be unrelated to retention performance, even during the first 1 min. of relearning. The function has zero slope until the loss in performance shown by the 24-mo. retention group.

3. The most important factor in retention is the *level* of proficiency achieved by the Ss during initial learning. This effect is shown to be just as important following long and short periods of no practice.

4. The *type* of initial training (amount of verbal guidance) is unrelated to retention performance when proficiency level after original learning is held constant.

5. Retraining administered under conditions of distributed practice proved to be superior to that administered under mass practice based upon a measure of performance during the final retraining session. However, on retesting 1 week later no difference was noted between the two retraining procedures. Thus, in terms of *transfer* to later performance there was no "permanent" disadvantage in massed retraining. Furthermore, both groups had improved beyond their original learning levels.

6. Predictions of individual differences in retention from independent ability measures were negligible. Retention appears more a function of specific task habits acquired, than of Ss' ability traits developed prior to training.

## REFERENCES

AMMONS, R. B., FARR, R. G., BLOCH, E., NEUMANN, E., DEY, M., MARION, R., & AMMONS, C. H. Long-term retention of perceptual-motor skills. *J. exp. Psychol.*, 1958, 55, 318–328.

BATTIG, W. F., NAGEL, E. H., VOSS, J. F., & BROGDEN, W. J. Transfer and retention of bidimensional compensatory tracking

after extended practice. *Amer. J. Psychol.*, 1957, 70, 75–80.

BELL, H. M. Retention of pursuit rotor skill after one year. *J. exp. Psychol.*, 1950, 40, 648–649.

BILODEAU, E. A., & BILODEAU, I. McD. Motor skill learning. *Annu. Rev. Psychol.*, 1961, 12, 243–280.

FLEISHMAN, E. A. Abilities and the learning of psychomotor skills. In P. H. Dubois, W. H. Manning, and C. J. Spies (Eds.), *Factor analysis and related techniques in the study of learning.* St. Louis: Washington University, 1959.

FLEISHMAN, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), *Training research and education.* Pittsburgh: Univer. Pittsburgh Press, 1962.

GAGNÉ, R. M., & FLEISHMAN, E. A. *Psychology and human performance.* New York: Holt, 1959.

JAHNKE, J. C. Retention in motor learning as a function of amount of practice and rest. *J. exp. Psychol.*, 1958, 55, 270–273.

JONES, E. I., & BILODEAU, E. A. Retention and relearning of a complex perceptual-motor skill after ten months of no practice. *HumRRO res. Bull.*, 1953, No. 53-15.

LEAVITT, H. J., & SCHLOSBERG, H. The retention of verbal and of motor skills. *J. exp. Psychol.*, 1944, 34, 404–417.

MENGELKOCH, R. F., ADAMS, J. A., & GAINER, C. A. The forgetting of instrument flying skills as a function of the level of initial proficiency. *USN Train. Dev. Cent. tech. Rep.*, 1958, No. 71-16-18.

PARKER, J. F., & FLEISHMAN, E. A. Prediction of advanced levels of proficiency in a complex tracking task. *USAF WADC tech. Rep.*, 1959, No. 59-255.

PARKER, J. F., & FLEISHMAN, E. A. Ability factors and component performance measures as predictors of complex tracking behavior. *Psychol. Monogr.*, 1960, 74(16, Whole No. 503).

PARKER, J. F., & FLEISHMAN, E. A. Use of analytical information concerning task requirements to increase the effectiveness of skill training. *J. appl. Psychol.*, 1961, 45, 295–302.

REYNOLDS, B., & BILODEAU, I. McD. Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, 44, 19–26.

# DISCRIMINATION OF THE REWARD IN LEARNING WITH PARTIAL AND CONTINUOUS REINFORCEMENT [1]

## STEWART H. HULSE

*Johns Hopkins University*

Over the past 20 years, psychologists have spent much time observing and manipulating the things that happen on nonreinforced trials in partial reinforcement learning situations. The result has been a massive array of information, and an equally massive array of theory, concerning the phenomena that partial reinforcement produces (Jenkins & Stanley, 1950; Lewis, 1960). The nonreinforced trial is the hallmark of partially reinforced learning, and the study of the stimulus and response correlates of nonreinforcement is, consequently, of unquestioned importance. But the nonreinforced trial is not the only place to look for information about the effects partial reinforcement produces.

It is quite conceivable that new information could be gained about partial reinforcement from a study of the behavior used to ingest the reward on reinforced trials. This notion is particularly appealing, first of all, in view of recent studies which have shown that consummatory responding, i.e., the rat's licking rate, varies quite systematically as a function of certain variables such as sweetness of reward, size of drop delivered from the drinking tube, and so on (Hulse & Bacon, 1962; Hulse, Snyder, & Bacon, 1960).

Consummatory behavior has another unique property in partial

reinforcement situations which makes its study potentially important. Practice of the learned response, running in an alley for example, is unavoidably confounded with number of reinforced and nonreinforced trials, but practice of the consummatory response in the goal box is not. This is true since the consummatory response cannot occur in overt and nonfractional form on nonreinforced trials. The reward is not there. Generally speaking, then, consummatory behavior takes place only as a function of number of reinforced trials. It follows that, while an examination of the development of the running response as a function of number of reinforcements is not logically justifiable, a similar examination of consummatory behavior is justifiable and might be quite interesting.

While the above is true in theory, in fact, some experiments (e.g., Marx, 1958) permit confounding of consummatory behavior and running in the sense that their procedures call for leaving a food cup, a drinking tube, or some other stimulus closely tied to the consummatory response, in the goal box on nonreinforced trials. Presumably, the rat makes abortive licks at the drinking tube, bites the food cup, or otherwise shows fractional components of the consummatory response. Other experimental procedures (e.g., Hulse, 1958; Weinstock, 1954) do, indeed, completely unconfound consummatory behavior from practice of the running response by deliberately removing

the food cup or drinking tube on nonreinforced trials. In both cases, the rationale is generally that of controlling secondary reinforcement or some other process presumed to transpire in the goal box. In neither case has an attempt been made to examine the specific behavior which the goal-box stimuli elicit on reinforced and nonreinforced trials.

The purpose of the present experiment was to examine consummatory behavior on both reinforced and nonreinforced trials as a function of three percentages of reinforcement of a running response in an alley. The confounding of consummatory behavior with practice of running in the alley was deliberately manipulated by making a drinking tube either available to $S$ or unavailable to $S$ on nonreinforced trials.

## METHOD

### Subjects and Apparatus

The $S$s were 50 male naive albino rats of the Sprague-Dawley strain obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The $S$s were 70 to 80 days of age at the time they started the experiment.

The apparatus was a 15-ft. U shaped enclosed runway. The sides of the U were 6 ft. long, and the base of the U was 3 ft. long. The start box was 7 in. long and 5½ in. wide, and the goal box was 20 in. long and 4 in. wide. The goal box was attached at right angles to the end of the runway such that $S$ made a left-hand turn into it. Inside height of the runway was 4 in. throughout. The runway was covered with hinged pieces of Plexiglas. The goal-box floor was covered with brass shim stock. The apparatus was painted flat black throughout.

Two guillotine doors separated the start and goal boxes from the alley. The goal-box door was located 4 in. before the turn into the goal box proper. Photocells, used to facilitate the timing of running behavior, were located 2 in. past each door.

Reinforcements were provided from a brass drinking tube (2 mm. inside diameter) located behind a Plexiglas shield at the far wall of the goal box. The tube was centered

in a ¾ × ⅜ in. vertical slot cut into the Plexiglas. A piece of Masonite could be placed over the entire far wall of the goal box such that $S$ could not see or reach the drinking tube on a particular trial.

Each lick on the drinking tube operated a pump system which delivered a drop of water of specified volume to the tip of the tube. The pump system, described in detail elsewhere (Hulse, 1960), consisted of an infusion pump operated by an electronic relay (Otis & Boenning, 1959). The electronic relay also operated a counter.

### Procedure

*Experimental design.*—Three percentages of reinforcement of the running response were used: 33%, 66%, and 100%. In addition, for the partial groups, the drinking tube was available to $S$ on nonreinforced trials (T) or it was not available to $S$ on nonreinforced trials (NT). An NT condition could not be included for the 100% $S$s, since this percentage of reinforcement required the presence of the drinking tube on all trials. Ten $S$s were used in each of the five groups called for by the design.

*Taming.*—On each of 10 taming days, the $S$s were handled freely and placed in groups of 5 or 6 into a large wooden box. Six water bottles were clipped to the outside of the box with their drinking tubes projecting through holes in the walls of the box. The $S$s were permitted to explore and to drink from the tubes for 15 min. This was the only water available during taming. Purina lab chow pellets were available in the individual home cages at all times.

*Training.*—Following taming, $S$s were given 60 training trials, 1 trial per day. On each trial, $S$ was placed in the start box, and after a 2- to 3-sec. delay, the start-box door was raised. The start-box and goal-box doors were lowered after $S$ had passed under them. After $S$ was removed from the goal box, it was returned to its home cage. Fifteen to 30 min. later, a water bottle was attached to the cage, and $S$ drank for 30 min. The $S$s were thus approximately 23 hr. thirsty at the time the trials began each day.

All $S$s were reinforced on each of the first 3 days of training. Reinforcement consisted of 600 licks on the drinking tube with the pump set to deliver .0053 cc of water with each lick. On Day 1, the drinking tube projected through the slot into the goal box. On Days 2 and 3, the tube was gradually withdrawn so that on Day 3, its tip was ⅛ in. behind the slot. The tube remained in this position for the rest of the experiment.

On Day 4, the first nonreinforced trial was introduced for the partial Ss. Thereafter, reinforced and nonreinforced trials were determined randomly according to the percentages of reinforcement called for by the experimental design. The only other restriction on randomization of reinforcement was that the last training trial was reinforced for all Ss.

On nonreinforced trials for the partial Ss tested under the T condition, the drinking tube was present in its usual location behind the slot, but the pump system was emptied of water. The S was thus free to lick from the tube, but no water was obtainable. On nonreinforced trials for Ss tested under the NT condition, the far wall of the goal box was covered with a piece of Masonite so that S could neither see nor lick from the tube. Goal-box confinement on nonreinforced trials was for 3 min.

*Measures of performance.*—Three measures of running behavior were recorded: start time, alley time, and goal-box time. Start time began when the start-box door went up and ended when S passed the photocell outside the start-box door. Alley time was the time S required to run from the first photocell to the second photocell at the goal-box door. Goal-box time was the time S required to run from the second photocell to the drinking tube and make the first lick. Goal-box times were not recorded on nonreinforced trials for the NT groups, since the drinking tube was not available. All times were recorded automatically on Standard Electric timers. The time scores were transformed to reciprocals and multiplied by 100 for purposes of the statistical analyses.

Licking times on reinforced trials were determined from a clock which started with the first lick on the drinking tube and stopped when S had completed its 600-lick allotment. These times were transformed to rates of licking, in licks per second. The number of licks Ss in the T groups emitted on nonreinforced trials was also recorded.

## RESULTS

*Licking.*—Figure 1 shows that if licking rates are plotted as a function of number of reinforcements, the rates for different percentages of reinforcement reach approximately the same asymptote at the last five reinforced trials. However, the figure also shows that the rate of increase
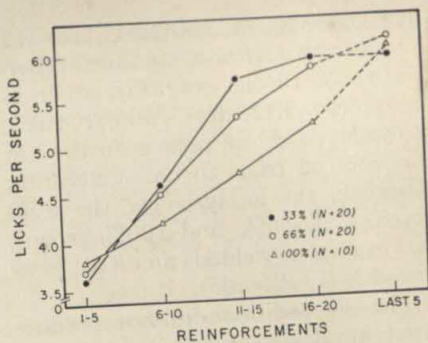


FIG. 1. The development of licking rates as a function of ordinal number of reinforced trials. (The scores have been combined for the partial groups, since the statistical analyses indicated the groups did not differ according to the T and NT variable.)

of licking over the first 20 reinforcements is quite different depending upon percentage of reinforcement. The curve for the 33% groups appears negatively accelerated, that for the 66% groups almost linear with a slight suggestion of negative acceleration, and that for the 100% group positively accelerated.

There is abundant statistical evidence to support the significance of the differences shown in Fig. 1. A simple $F$ test for percentage of reinforcement based on means for the last 10 of the first 20 trials shown in the figure yields an $F$ of 7.92 ($df = 2/47$, $P < .01$). An analysis of variance of 5-trial means for the first 20 reinforcements for the 33% and 66% T groups and the 100% group, i.e., for those groups which had comparable goal-box conditions on all training trials, yields an $F$ for percentage of 4.44 ($df = 2/27$, $P < .05$). This analysis also shows a significant Percentage × Blocks of Trials interaction ($F = 2.50$, $df = 6/81$, $P < .05$) which indicates that the three groups are progressing towards their final asymptote at different rates. A similar analysis based on 5-trial means

for the 33% and 66% T and NT groups also shows a significant Percentage × Blocks of Trials interaction ($F=3.12$, $df=3/108$, $P<.05$). No analysis shows means on the last 5 reinforced trials to be significantly different; the analysis for the combined 33%, 66%, and 100% groups, for example, yielded an $F$ of 3.00 ($df = 2/47$, $P > .05$).

The T condition did not produce different licking behavior on reinforced trials than the NT condition. The analysis of variance of 5-trial means for the 33% and 66% T and NT groups provides no evidence that licking rates differ as a function of the availability of the tube in the goal box on nonreinforced trials ($F=1.66$, $df=1/36$, $P>.05$). This fact is equally true for the last 5 reinforced trials of the 60 training trials ($P > .05$). Moreover, the T and NT condition does not interact with percentage of reinforcement ($P > .05$).

Figure 2 shows that the 66% T group emits more licks on nonreinforced trials than the 33% T group ($F = 6.02$, $df = 1/18$, $P < .05$), number of licks decreases for both groups across blocks of nonreinforced trials ($F = 10.92$, $df = 3/57$, $P < .01$), but the difference in number of licks between the groups across blocks of trials does not change in size ($P > .05$). Of particular importance is the fact that the absolute number of licks for both groups is small, and most of the decrease in number of licks occurs during the course of the first 5 to 10 nonreinforced trials.
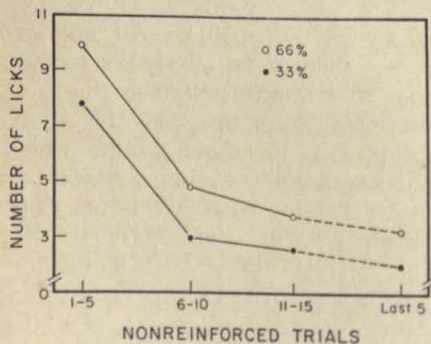
Many of the $S$s in the T groups failed to lick at all on some nonreinforced trials. This happened more frequently for the 33% T group than for the 66% T group where, out of the 200 trials represented by the first 15 and last 5 nonreinforced trials, the frequencies are 73 and 60 nonlick trials, respectively. A Wilcoxon $T$ test shows this difference to be significant ($P < .05$).

*Running.*—We noted that separate analysis of running behavior for reinforced and nonreinforced trials is not logically justifiable, since practice of running is confounded with the number of such trials. Strictly speaking, however, this is not true for goal-box speeds, since this measure reflects behavior which occurs after $S$s enter the goal box and expose themselves directly to the different stimulus conditions correlated with reinforcement and nonreinforcement.

Although an analysis of variance of blocks of 5 trials for the first 20 reinforced trials shows a significant Blocks × Percentage interaction ($F = 2.64$, $df = 6/81$, $P < .05$), goal-box speeds differ only for the first 5 reinforced trials. For this block of trials, speeds are greater the higher the percentage of reinforcement. The groups do not differ on the last 5 reinforced trials of the 60 training trials ($P > .05$).

An analysis of variance of goal-box speeds on the first five and last five nonreinforced trials for the 33% and 66% T groups shows that the 66% group ran to the tube faster than



FIG. 2. Number of licks on nonreinforced trials for the 33% and 66% T groups as a function of ordinal number of nonreinforced trials.

the 33% group ($F = 7.47$, $df = 1/18$, $P < .05$), but that goal-box speeds for the two groups *decrease* from the beginning to the end of training ($F = 6.70$, $df = 1/18$, $P < .05$). The interaction between the variables is not significant.

The differences obtained for behavior in the goal box are indeed unique to the goal box and did not result from some extraneous factor such as differential handling by $E$ at the beginning of reinforced as compared with nonreinforced trials. This is substantiated by a control analysis run on alley speeds for Reinforced Trials 16 to 20 which showed a significant effect for percentage of reinforcement ($F = 4.78$, $df = 2/47$, $P < .05$). Partially reinforced $S$s run faster than continuously reinforced $S$s, as we would expect, since the former have had many more trials in the alley.

An analysis of start and alley speeds at the end of 60 training trials shows little except that a 15-ft. U shaped alley produces a great deal of response variability. Analyses of variance based on means for the last 20 training trials for start speeds show no significant differences due to any of the variables. The same is true for alley speeds, except for an analysis which compared means for the 33% and 66% condition and the T and NT condition. Here, a significant percentage effect was obtained ($F = 4.27$, $df = 1/36$, $P < .05$) which indicated that 66% reinforcement produced higher speeds than 33% reinforcement. This effect was not significant, however, for an analysis based on means from the 100%, 66% T, and 33% T groups.

## DISCUSSION

Rats quickly learn to detect whether or not the drinking tube will produce water, and their licking rates rapidly increase as a function of number of reinforcements. However, the *rate* with which licking rates increase over reinforced trials is a function of the number of intervening nonreinforced trials.

We can account for this phenomenon on the assumption that a stimulus-discrimination process takes place in the goal box for the partial groups. If we think of the drinking tube and fluid as discriminative stimuli for approaching the tube and drinking, the data suggest that rats quickly learn to attend to these stimuli and to do the appropriate thing when they are present or absent.

First, licking on nonreinforced trials occurs, but it rapidly extinguishes. The $S$s in the T condition learned to take only a lick or two on the dry drinking tube on nonreinforced trials, and they learned to do this early in training over the first 5 to 10 nonreinforced trials. Also, after a given number of nonreinforced trials, the $S$s in the 66% T group have had more reinforced trials than $S$s in the 33% T group and have, presumably, developed a stronger consummatory response. We might, therefore, expect a greater tendency for them to generalize licking on reinforced trials to licking on nonreinforced trials. This apparently occurred, since the 66% T group emitted more licks on nonreinforced trials than the 33% T group. Finally, in this connection, $S$s in the partial groups were apparently so set to discriminate cues associated with reinforcement from cues associated with nonreinforcement that they were able to outwit $E$ in their ability to identify a nonfunctional drinking tube. Under the T condition, all $S$s failed to lick on some nonreinforced trials. Possibly, since the pump system was drained on nonreinforced trials, the $S$s could discriminate after they entered the goal box whether or not water was visible at the opening of the tube. Possibly, for the same reason, they discriminated an odor difference in the goal box; given the chemical content of local laboratory tap water, this is not an inconceivable proposition.

Second, goal-box speeds on nonreinforced trials decrease as a function of the number of such trials. Since the licking data show that partial $S$s could sometimes detect a dry drinking tube before they took a lick, approaching the drinking tube on a nonreinforced trial extinguished to some extent. Moreover, the effects of this process appear to have generalized to suppress the speed with which partial $S$s approached the drinking tube on reinforced trials. Thus, goal-box speeds on reinforced trials increase at approximately the same rate regardless of percentage of reinforcement. This is to be contrasted with the usual finding, obtained for alley speeds in the present experiment, that increments in response strength as a function of number of reinforced trials will be much greater the lower the percentage of reinforcement of the response.

Finally, the consummatory response develops over reinforced trials as $S$s learn to discriminate the special significance of the drinking tube and fluid as stimuli for consummatory responding. For a fixed number of reinforcements, the rate with which this discrimination develops will increase as the number of intervening nonreinforcements increases. We would therefore expect the 33% groups to learn the consummatory response faster than the 66% groups. Since the 100% group received no discrimination training with respect to the drinking tube, we would expect this group to develop the consummatory response slowest of all. The data clearly support these conclusions.

It seems clear that partial reinforcement provides discrimination training for reward stimuli in the goal box, but continuous reinforcement does not. During partial reinforcement, in effect, behavior is critically focused on the reward and its stimulus properties because of the contrast in goal-box conditions on reinforced as compared with nonreinforced trials. It follows that, if partial reinforcement is used to condition a response, the development of response strength may be more critically determined by stimuli correlated with the

reward, such as its sweetness, than if continuous reinforcement is used. We might expect, for example, that after partial reinforcement resistance to extinction would increase as a function of the sweetness of a reward. After continuous reinforcement, however, there should be much less correlation between reward sweetness and resistance to extinction. Hulse and Bacon (1962) showed these predictions to hold following training with different concentrations of saccharin. Hulse (1958) showed much the same thing for different-sized food rewards.

Some additional support for these hypotheses comes from another source. Jenkins (1961) found that resistance to extinction in the presence of a stimulus was greater if that stimulus had been paired with reinforcement during discrimination training as compared with continuous reinforcement training. Jenkins' stimulus was a light pattern projected on $S$'s response key. The analogous stimulus in the present approach is the reinforcing stimulus itself, and this is not presented, of course, during extinction. If the effect Jenkins noted is to occur in straightforward fashion, it would have to operate through some mechanism other than the reward stimulus per se. There is a clear parallel between Jenkins' approach and the approach outlined here, and the correspondence between Jenkins' data and the data discussed here is interesting. This apparent correspondence must remain suggestive for the present, however.

## SUMMARY

Fifty albino rats were given 60 training trials in a 15-ft. U shaped alley. They were reinforced with 600 licks of water from a drinking tube on 33%, 66%, or 100% of the trials. Half the partial $S$s could lick on the dry drinking tube on nonreinforced trials; for the other half, the tube was blocked such that the $S$s could neither see it nor lick it. Start speeds, alley speeds, and goal-box speeds, licking rates on reinforced trials, and number of licks on nonreinforced trials were recorded.

The results show that the partial groups

developed licking rates faster in the goal box, as a function of number of reinforcements, than the continuous group. This happened whether or not the tube was available on nonreinforced trials. Licking rapidly extinguished, and goal-box speeds decreased, on nonreinforced trials for the partial groups. Start speeds and alley speeds did not vary as a function of any of the experimental variables; this may have been due to excessive response variability produced by the very long runway.

The data suggest that partial reinforcement produces a very powerful discrimination of reward stimuli in the goal box. This process may be a factor in experiments which have shown that stimulus variables correlated with the reward, such as its sweetness or size, have different effects if they are used with partial as compared with continuous reinforcement.

## REFERENCES

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HULSE, S. H. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

HULSE, S. H., & BACON, W. E. Supplementary report: Partial reinforcement and amount of reinforcement as determinants of instrumental licking rates. *J. exp. Psychol.*, 1962, **63**, 214–215.

HULSE, S. H., SNYDER, H. L., & BACON, W. E. Instrumental licking behavior as a function of schedule, volume, and concentration of a saccharine reinforcer. *J. exp. Psychol.*, 1960, **60**, 359–364.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121.

JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

MARX, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

OTIS, L. S., & BOENNING, R. A. A transistorized circuit for recording contact responses. *J. exp. Anal. Behav.*, 1959, **2**, 289–291.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

# MEDIATED ASSOCIATION IN A PAIRED-ASSOCIATE TRANSFER TASK

DAVID S. PALERMO

*Institute of Child Development, University of Minnesota*

A reliable condition for producing negative transfer in paired-associate learning occurs in the A-B, A-C paradigm. In this paradigm, one set of responses is learned to a set of stimuli and subsequently, new responses are learned to the same stimuli. Another approach to an understanding of the learning of successive lists of paired associates has used an A-B, B-C, A-C paradigm for an experimental group and an A-B, D-C, A-C paradigm for a control group. In such mediated association studies (e.g., Norcross & Spiker, 1958), it has been assumed that the A-C list is learned more rapidly by the experimental group because the learning of the B-C list provides a mediating link between A and C of the following nature:

$$S_A \qquad\qquad S_C$$
$$\downarrow \searrow \qquad\qquad \downarrow$$
$$R_A s_a \longrightarrow R_B s_b \longrightarrow R_C$$
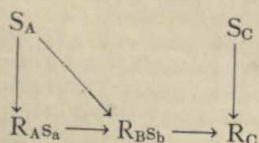
The $s_b \rightarrow R_C$ association learned in the B-C list provides the associative or mediational link assumed to facilitate the learning of the A-C list.

The present experiment was designed to determine whether mediated association effects can be demonstrated within an A-B, A-C paradigm. The usual paradigm was modified to the extent of presenting lists in which each response was paired with two stimuli rather than one response with each stimulus. Therefore, there were only half as many different responses as stimuli in both the A-B and the

A-C lists. Upon reaching criterion on the A-B list, each response and its associated stimulation had been conditioned to two stimuli of the list. By appropriate pairing of the new responses with the old stimuli in a "mixed" A-C list (Twedt & Underwood, 1959), it was possible to arrange stimuli and responses which, according to the mediational hypothesis, would lead to facilitation or impairment of parts of the A-C list *during* the learning of the list. Thus, the learning of a particular pair in the A-C list was assumed to mediate responses which would affect the learning of another pair within the list.

## METHOD

*Apparatus.*—The apparatus was a modified Card Master. Basically, it consisted of a box 13 × 13 × 9½ in. The front face of the apparatus contained a 3 × 6 in. aperture covered by clear plastic. A mechanical arrangement delivered to the aperture a 3½ × 6 in. plastic card from the bottom of a stack of cards. Mounted on each card was a pair of Stanford-Binet Picture Vocabulary pictures reproduced by a Thermofax process for this purpose. Gray metal doors served as shutters to expose independently the left and right pictures on the cards. Following a presentation, the doors closed simultaneously and the card was released from the aperture and returned to the top of the stack by a conveyor belt. A system of electronic timers controlled the rate of presenting the cards and the exposure times of the stimulus and response picture.

*Experimental design.*—All Ss learned two lists, each of which was composed of six stimulus pictures and three response pictures paired so that each response was learned to two stimuli. Table 1 presents the design of the experiment, with the picture names, which provides that each S serve under each

## TABLE 1

### DESIGN OF THE EXPERIMENT

| List I | | List II | | Cond. |
|---|---|---|---|---|
| S | R | S | R | |
| $S_1$ STOOL | $R_A$ SCISSORS | $S_1$ STOOL | $R_D$ HOUSE | I |
| $S_2$ SHOE | $R_A$ SCISSORS | $S_2$ SHOE | $R_D$ HOUSE | I |
| $S_3$ GLASSES | $R_B$ BED | $S_3$ GLASSES | $R_E$ BASKET | II |
| $S_4$ CUP | $R_B$ BED | $S_4$ CUP | $R_F$ KNIFE | II |
| $S_5$ CLOCK | $R_C$ HAND | $S_7$ FORK | $R_E$ BASKET | III |
| $S_6$ TABLE | $R_C$ HAND | $S_8$ TREE | $R_F$ KNIFE | III |

of three experimental conditions. For two pairs in List 2 the mediational chain, established during the learning of the pairs, is expected to facilitate learning (Cond. I). For two other pairs the mediational link is expected to elicit incorrect responses and, thus, interfere with learning (Cond. II). The other two pairs involve two new stimuli as well as the two new responses and thus have no experimentally established mediating tendencies (Cond. III). In the case of Cond. I and II, it is assumed that when a stimulus is presented in List 2, the List 1 response and attendant stimulation will occur either overtly or covertly. It is also assumed that the response of List 2 will be conditioned to the attendant stimulation of the List 1 response.

For example, in List 2 of Table 1 the learning of the pair STOOL-HOUSE ($S_1 \rightarrow R_D$) in Cond. I is expected, through the establishment of the association SCISSORS-HOUSE ($s_a \rightarrow R_D$), to increase the tendency to make the response HOUSE ($R_D$) to SHOE ($S_2$). Similarly, the learning of SHOE-HOUSE ($S_2 \rightarrow R_D$) is expected to facilitate the learning of STOOL-HOUSE ($S_1 \rightarrow R_D$). In the case of Cond. II, it is expected that the learning of GLASSES-BASKET ($S_3 \rightarrow R_E$), in List 2, will interfere with the learning of CUP-KNIFE ($S_4 \rightarrow R_F$) because of the establishment of the association of bed-BASKET ($s_b \rightarrow R_E$). Similarly, the establishment of the association bed-KNIFE ($s_b \rightarrow R_F$) during the learning of CUP-KNIFE ($S_4 \rightarrow R_F$) will interfere with the learning of GLASSES-BASKET ($S_3 \rightarrow R_E$). The occurrence of BASKET ($R_E$) to CUP ($S_4$) or KNIFE ($R_F$) to GLASSES ($S_3$) would be considered a mediated error. In the case of Cond. III, the stimuli in List 2 have no experimentally established mediating tendencies since they have not been previously presented.

*Subjects.*—The Ss were drawn from Grades 3–6 ($N$s = 12, 12, 15, and 18, respectively) of the University of Minnesota

Elementary School.[1] All Ss were naive with respect to verbal learning except the Grade 6 children who had participated in a paired-associate learning experiment 1 yr. previously. Seven Ss had to be discarded due to apparatus failure and 19 were discarded because they failed to reach criterion on List 1 within 27 trials (11, 4, 4, and 1 in Grades 3–6, respectively). All groups received the same form of List 1. One-third of the Ss in each grade level were randomly assigned to each of three groups designed to control for possible differential difficulty of the stimulus-response pairs in List 2. The groups differed with respect to which of the three forms of List 2 they received. The three forms were constructed so that a given set of two List 2 pairs was used in each of the three conditions.

*Procedure.*—The Ss were given two experimental sessions separated by 1–4 days. Each S was brought by E from the classroom to the experimental room, and seated facing a 37 × 49 in. panel with a 3 × 6 in. aperture through which S viewed the stimulus cards. A shaded 40-w. light directly above the aperture provided illumination. The apparatus was located behind the panel in an adjoining room. On the table was a Webster Teletalk two-way speaker. The E gave the instructions through the speaker system and recorded the verbal responses as they were given by S. Thus, S was alone during both experimental sessions.

At Session 1 instructions explaining the task were read to S and a pair of pictures, not subsequently used in the experiment, were shown to familiarize him with the apparatus

and procedure. The $S$ was not required to name the stimulus picture but was asked to anticipate the response picture by naming it. No $S$ had any difficulty in giving the appropriate name to the stimuli. List 1 was presented immediately and continued until $S$ reached a criterion of three successive errorless trials.

At Session 2 each $S$ relearned List 1 to three successive errorless trials and then was told that another list would be presented. After approximately 2 min., necessary to change lists, List 2 was presented for nine trials or three successive errorless trials, whichever was fewer. If the criterion was reached prior to the nine trials, it was assumed for the analysis that no additional errors would have been made.

The pictures in List 1 were randomly designated as stimuli and responses and randomly paired together with the restriction that no obviously highly associated pictures were paired together. All lists were presented in three varying orders to control for serial learning. The orders were randomly determined with the exception that no response picture was presented twice in succession. A 3-sec. anticipation period, a 3-sec. joint presentation of stimulus and response, and a 3-sec. interval between pairs was used throughout the experiment. All verbal responses occurring in the anticipation interval were recorded verbatim throughout the experiment.

RESULTS

Table 2 presents the means and $SD$s for the number of correct anticipations in each condition for each grade. An evaluation of the differ-

ences among these groups employed an analysis of variance in which the main effects of Control List Groups and of Grade were between-$S$ factors while the main effect of Conditions was an intra-$S$ factor (Lindquist, 1953, p. 281, Type III). The means for the groups designed to control for possible differential difficulty of the picture pairs did not differ significantly ($F = 1.68$, $df = 2/45$, Error$_b = 14.54$), nor did the four grades ($F = 2.56$, $df = 3/45$, $.05 < P < .10$). The means for the three experimental conditions differed beyond the .001 level ($F = 14.88$, $df = 2/90$, Error$_w = 3.32$). Individual $t$ tests for related measures showed that the differences between Cond. I and III and between Cond. II and III were significant ($P < .01$), but the difference between Cond. I and II was not ($.05 < P < .10$). None of the interactions was significant.

Although the analysis of variance indicated no significant interaction between Grade and Conditions, it may be seen in Table 2 that there appears to be an interaction between Cond. I and II and grade level. Performance of $S$s in Grades 3 and 4 shows clearly the effects of the two conditions, but for Grade 5 the effect is much less pronounced and for

TABLE 2

MEANS AND $SD$s OF TRIALS TO LEARN LIST 1 AND FOR CORRECT RESPONSES ON LIST 2 UNDER EACH CONDITION FOR EACH GRADE

| Grade | N | List 1 | | List 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Trials to Learn | | Cond. I | | Cond. II | | Cond. III | | Cond. I, II, III | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 3 | 12 | 19.42 | 7.64 | 11.58 | 3.10 | 10.33 | 3.40 | 12.50 | 2.81 | 11.47 | 3.24 |
| 4 | 12 | 19.42 | 7.27 | 12.92 | 2.50 | 10.83 | 2.11 | 14.33 | 2.28 | 12.69 | 2.72 |
| 5 | 15 | 19.27 | 6.51 | 12.61 | 3.26 | 12.27 | 2.52 | 13.27 | 2.77 | 12.71 | 2.89 |
| 6 | 18 | 12.39 | 3.27 | 13.11 | 2.08 | 13.44 | 2.09 | 14.67 | 1.20 | 13.74 | 1.95 |
| Total | 57 | 17.16 | 6.96 | 12.61 | 2.80 | 11.93 | 2.82 | 13.77 | 2.44 | 12.77 | 2.80 |

Grade 6 the means are actually in the opposite direction of that predicted. Consequently, an analysis of variance was conducted on the data for Cond. I and II alone to permit a more sensitive test for the interaction indicated in Table 2. The difference between the conditions was significant at the .05 level ($F = 4.14$, $df = 1/45$, $Error_w = 3.22$), but the Grade $\times$ Conditions interaction was not significant by usual standards ($F = 2.48$, $df = 3/45$, $.05 < P < .10$).

An analysis of the errors made in response to stimulus members of the pairs in Cond. II and III was made by the procedure used by Norcross and Spiker (1958). The mean number of reversals per $S$ for Cond. II was 1.33 and for Cond. III, .93 ($P = .10$, related $t$ test). For Grades 3 and 4, where the mediation effect was more apparent, there were 1.58 reversals in Cond. II and 1.00 in Cond. III ($P = .05$, related $t$ test). However, these suggestions of differences must be discounted because the ratio of the number of such intrusion errors to total errors, was approximately equal in each comparison of Cond. II and III.

## DISCUSSION

These data provide evidence that the conditions designed to produce mediated associations result in differential effects upon the amount of transfer in an A-B, A-C transfer paradigm. Performance on the pairs assumed to be facilitated by the mediation of correct responses was superior to performance on pairs in which interference was assumed due to the mediation of incorrect responses. This result had been found in a preliminary study using 22 Grade 6 $S$s in which List 2 was composed of four pairs designed to produce facilitation through mediation (Cond. I) and four pairs designed to produce interference through mediation (Cond. II) but without the

control pairs (Cond. III). The mean number of correct responses for Cond. I and II in that study was 25.00 ($SD = 4.00$) and 23.32 ($SD = 3.76$), respectively, yielding a $t$ which was significant at the .025 level.

While the relationship between the conditions assumed to produce facilitation and interference due to mediation seems to be a stable one under these conditions, some caution should be made in generalizing these results. Attempts to replicate the findings of this study with college students and Grade 10 children using the same design failed to yield significant differences among the three conditions. In the study with college students, low-association nonsense syllables were used as stimuli and responses, and in the Grade 10 study high-frequency adjectives were used.

The suggestion of an interaction between grade level and the effects of Cond. I and II in this study along with the failure to replicate the findings with college students and tenth graders might indicate that some factor or factors related to age may be important.

In the present study, performance on the control pairs was superior to performance on the pairs assumed to be facilitated by mediation as well as on the pairs assumed to be impaired by mediation. These results are in contrast with those of Norcross and Spiker (1958). Performance on their control pairs was superior to performance on pairs in which interference was expected and inferior to performance on pairs in which facilitation was expected in a three-list paradigm. Apparently, the strong associative interference effects of the A-B, A-C paradigm in the present study outweigh any facilitation introduced by the mediated associations.

Ignoring the mediation effects, this study replicates the findings of the Spiker and Holton (1958) study in which it was demonstrated, using a motor paired-associate task, that learning of an A-B, A-C series results in associative interference relative to an A-B, D-C series. In the present experiment the control pairs are comparable to the D-C pairs

of the Spiker and Holton study and were learned significantly faster than either the Cond. I or II pairs.

## SUMMARY

An experiment was conducted to study facilitation and impairment of performance as a function of mediated associations in a modified A-B, A-C transfer paradigm. Children in Grades 3, 4, 5, and 6 were required to learn two lists of six paired associates composed of six stimuli and three responses. In List 2, the pairs were arranged so that learning could be facilitated by mediated associations or impaired by mediated associations. In addition, control pairs, in which no experimentally manipulated mediation was present, were included.

The results indicated that the condition designed to produce facilitation through mediated associations led to superior performance when compared with the condition designed to produce impairment through mediated associations. Performance on the control pairs was superior, however, to that of both of the mediation conditions. The results were discussed in terms of age related variables and the associative interference effects of the A-B, A-C paradigm.

## REFERENCES

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

NORCROSS, K. J., & SPIKER, C. C. The effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, 55, 129–134.

SPIKER, C. C., & HOLTON, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.*, 1958, 56, 123–132.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, 58, 111–116.

# PREDICTION OF PREFERENCE, TRANSPOSITION, AND TRANSPOSITION-REVERSAL FROM THE GENERALIZATION GRADIENT [1]

WERNER K. HONIG

*Denison University*

The present study concerns the prediction of differential responding in an operant situation from the generalization gradient. Gradients were obtained on the spectral continuum with the technique of Guttman and Kalish (1956), while preference tests involving the generalization stimuli were concurrently administered. In a general sense, this is an attempt to predict choice behavior when the responsiveness to each single stimulus is known. More specifically, it involves a test of the power of the generalization gradient in the determination of behavior in situations more complex than the single-stimulus condition in which it is assessed.

The desired relationships were investigated after simultaneous and successive[2] discrimination training as well as simple acquisition, in order

to determine whether the peak shift obtained by Hanson (1959) for the postdiscrimination gradient (PDG) is accompanied by transposition and transposition-reversal as predicted by Spence (1937). In previous transposition studies (e.g., Baker & Lawrence, 1951; Ehrenfreund, 1952; Kendler, 1950) the PDGs have not been obtained; it could not be determined, therefore, whether the consistent failure to obtain clear-cut transposition-reversal in such studies was due to the form of the PDGs or to a lack of correspondence between them and stimulus preferences on the transposition tests. In the present study, PDGs and transposition tests were obtained concurrently.

## EXPERIMENTS 1 AND 2

### Method

*Apparatus.*—The automatic key pecking apparatus used in the present investigation was similar to that employed in other spectral generalization studies (e.g., Guttman & Kalish, 1956) except that it included two separately illuminated keys $7\frac{1}{2}$ in. from the floor with 2 in. between centers. Bausch and Lomb monochromatic interference filters, with band widths at half height of 7–9 m$\mu$, provided the different spectral values. In the text, stimuli will be referred to by the nominal value of the filters, but the spacing of points on the abscissae of the figures corresponds to the actual transmission peaks which differ slightly in some cases. A K-2 yellow filter was inserted when necessary to prevent the transmission of the visible second-order spectrum produced by the filters. The various spectral values were equated for apparent brightness by a human observer with the aid of a Macbeth illuminometer. This appears justified by the work of

[2] A *successive discrimination* refers in this paper to the case where only the positive or the negative stimulus value is presented at one time to *S*, or the "go, no-go" discrimination. The term has been used in this sense by Grice (1949) and Baker and Lawrence (1951). Other authors (Bitterman, Spence) have more recently used *successive discrimination* to refer to a conditional left-right discrimination, but this usage is not intended here.

## TABLE 1

COMPOSITION OF STIMULUS PAIRS ON THE
GENERALIZATION TESTS IN EXP. 1 AND 2

| Experiment 1 | | Experiment 2 | |
|---|---|---|---|
| Absolute Value (m$\mu$) | Difference from CS (m$\mu$) | Absolute Value (m$\mu$) | Difference from CS (m$\mu$) |
| 490, 510 | −60, −40 | 490, 590 | −60, +40 |
| 510, 520 | −40, −30 | 590, 520 | +40, −30 |
| 520, 530 | −30, −20 | 520, 570 | −30, +20 |
| 530, 540 | −20, −10 | 570, 540 | +20, −10 |
| 540, 550 | −10, 0 | 540, 550 | −10, 0 |
| 550, 560 | 0, +10 | 550, 560 | 0, +10 |
| 560, 570 | +10, +20 | 560, 530 | +10, −20 |
| 570, 580 | +20, +30 | 530, 580 | −20, +30 |
| 580, 590 | +30, +40 | 580, 510 | +30, −40 |
| 590, 610 | +40, +60 | 510, 610 | −40, +60 |

Blough (1957), which indicates that the spectral luminosity function of the pigeon is similar to that of man. No illumination was provided in the box other than the light falling on the keys, except during the presentation of grain, when a magazine light went on.

*Subjects.*—The Ss were 38 white Carneau pigeons reduced to 75% of their free-feeding weight. Twenty-two experimentally naive animals were used in Exp. 1. The 6 Ss used in Exp. 2 had served in a previous experiment in which 10 sessions of pecking at a 550-m$\mu$ stimulus on a variable interval schedule were followed by a generalization test in which the values used in the present study were presented. The previous experiment was run in a different apparatus which had one key.

*Procedure.*—Over several daily sessions, S was magazine trained, conditioned to peck at a 550-m$\mu$ key, and given 60 continuous reinforcements. Following this, S received 10 daily sessions of variable interval (VI) training with a mean interreinforcement interval of about 1 min. Each session consisted of 30 1-min. periods separated by 10 sec. of blackout. During half the periods, one key was illuminated by 550 m$\mu$. During the remaining periods, both keys were illuminated by 550 m$\mu$. The single stimulus appeared for half the time on the right and for half the time on the left key. The appearance of the single stimulus, and the availability of reinforcement in the two-stimulus case, were balanced between left and right keys. The order of conditions was randomized. Five of the 22 Ss in Exp. 1 received training beyond

10 sessions until they reached a criterion of 900 responses in a single session.

This training procedure was abbreviated for Exp. 2, as magazine training and conditioning of key pecking could be omitted. After 10 continuous reinforcements, VI training began immediately in the manner described above. All Ss received 7 rather than 10 sessions of VI training, and easily reached 900 responses per session by the seventh session.

On the 2 days following the last session of VI training, S received a generalization test under extinction, consisting of 189 30-sec. periods of stimulus presentation alternating with 10 sec. of blackout. Six blocks of 21 periods each were given on the first day of testing and three on the second. A single stimulus value was presented on 11 of the 21 periods within each block and a pair of stimulus values on the remaining 10 periods. The single stimulus values ranged from 490 to 610 m$\mu$ in 10-m$\mu$ steps, with the omission of 500 and 600 m$\mu$. The stimulus pairs were composed of these values as summarized in Table 1. In Exp. 1, the pairs were composed of adjacent stimuli in the series. In Exp. 2, they were so chosen that the differences in m$\mu$ between the test values and the CS in each pair correspond to those for Exp. 1 in size, but have opposite signs. In other words, in Exp. 1 the stimuli in each pair lie on the "same side" of the CS, while in Exp. 2 they lie on "opposite sides."

The 21 stimulus conditions were randomized within each block of presentations. Within a given block, 490, 520, 540, 560, 580, and 610 m$\mu$ were presented on one key and 510, 530, 550, 570, and 590 m$\mu$ were presented on the other key. This arrangement alternated on successive blocks.

By means of this testing procedure, it was possible to present a generalization test consisting of singly presented stimulus values in the same session with a test consisting of different pairs of values. These will be distinguished as single-stimulus (SS) and double-stimulus (DS) generalization tests, even though they occurred in the same session. The testing procedure provided control for differences between animals and for the effects of extinction in the course of testing.

## Results

*Experiment 1.*—The data obtained from the generalization tests are presented in Fig. 1. The mean total responses to SS values are shown by

the single-stimulus gradient (SSG); the mean total responses to the members of each stimulus pair are shown by the adjacent bars. The filled bar in each pair represents the stimulus nearer the CS. When the responses to each stimulus *value* on the DS test are summed across the two pairs in which that value appears, the double-stimulus gradient (DSG) is obtained.

The SSG peaks at 550 mµ and decreases to both sides of the CS. There are inversions between 490 and 510 mµ and between 560 and 570 mµ. The former inversion is small and may be due to random error, as the gradient is almost flat in that region. The latter inversion is sometimes found when stimuli in this region are equated for brightness (Honig, Thomas, & Guttman, 1959).

The direction of preference within each stimulus pair agrees with the difference between the number of responses given to the members of that pair in the SS test. Accordingly, the stimulus nearer 550 mµ, received more responses in all pairs except 490, 510 mµ and 560, 570 mµ. It also appears that the degree of preference is systematically related to the difference in response level indicated for the corresponding stimuli on the SSG.

The DSG is very similar to the SSG except that the central values are somewhat higher. The "predicted DSG values" are derived from a method of predicting response totals in the DS situation which will be presented below.

*Experiment 2.*—The SSG and DSG are so similar to those obtained in
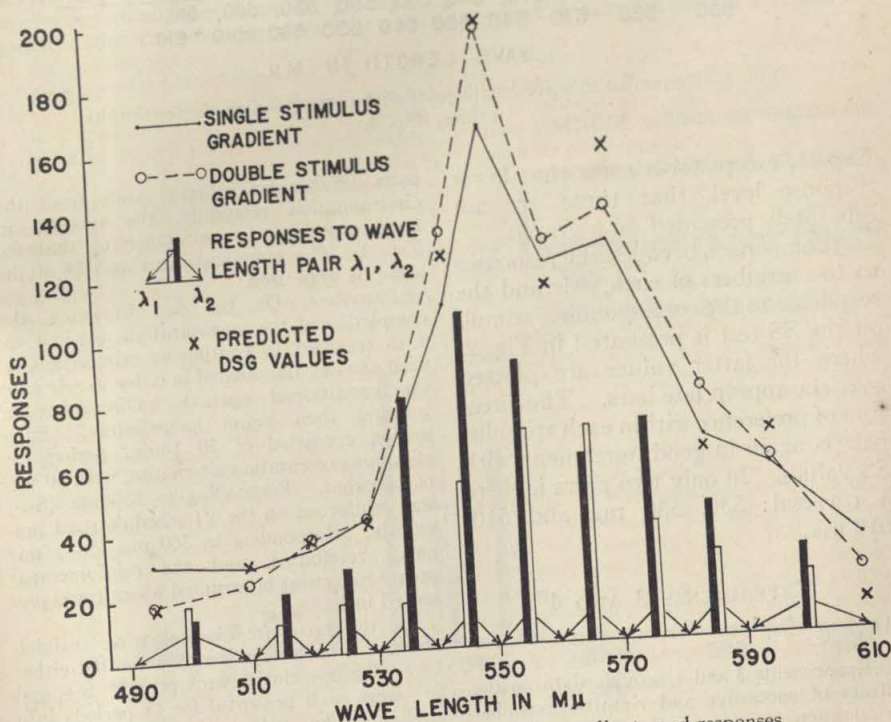


FIG. 1. Single-stimulus and double-stimulus gradients and responses to wave-length pairs, Exp. 1.
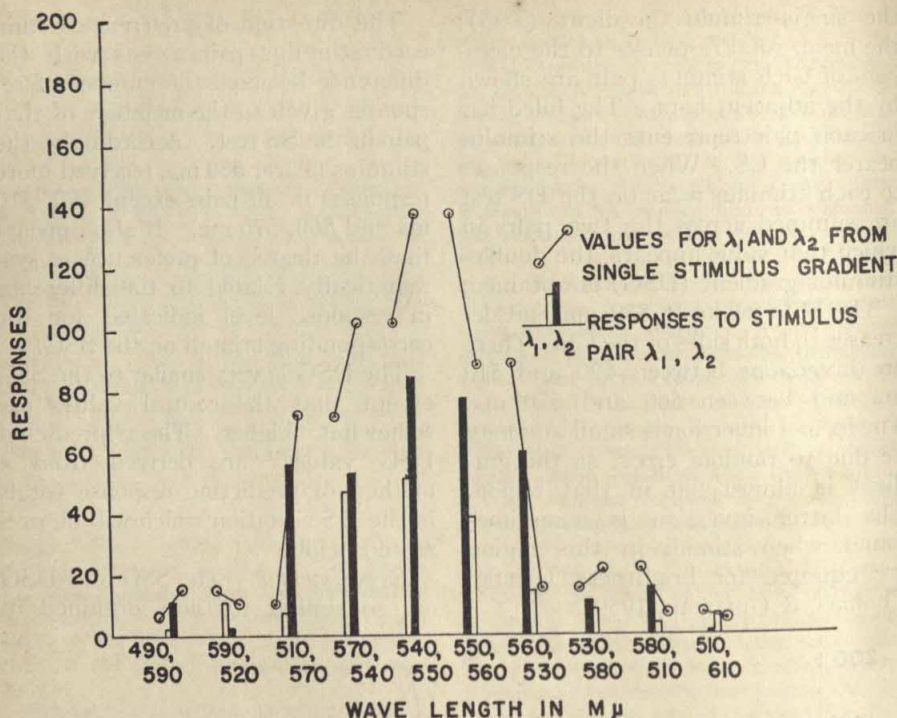
FIG. 2. Responses to wave-length pairs with corresponding single-stimulus
values, Exp. 2.

Exp. 1, except for a somewhat lower response level, that these are not separately presented on a graph.

A comparison between the responses to the members of each pair and the responses to the corresponding stimuli on the SS test is presented in Fig. 2, where the latter values are plotted over the appropriate bars. The direction of preference within each stimulus pair is again in good agreement with SS values. In only two pairs is there a reversal: 530, 580 m$\mu$ and 510, 610 m$\mu$.

## EXPERIMENTS 3 AND 4

### Method

Experiments 3 and 4 provide data on the effects of successive and simultaneous discrimination training on the SSG and DSG and on the distribution of responses between pairs of stimulus values. Aside from the discrimination procedure, the method in these experiments was similar to that for Exp. 1. The same apparatus and 14 of the same Ss were used.

*Procedure.*—On the day following the completion of the generalization test in Exp. 1, S received 10 continuous reinforcements with one key illuminated in order to reinstate the conditioned operant. Discrimination training then began immediately. Each session consisted of 30 1-min. periods of stimulus presentation alternating with 10 sec. of blackout. Responding to 550 m$\mu$ (S+) was reinforced on the VI schedule used previously. Responding to 560 m$\mu$ (S−) was never reinforced, and the reinforcement programmer was interrupted when it was presented in Exp. 3.

In the *successive* discrimination training of Exp. 3, both keys were illuminated either by S+ or S− during each period. S+ and S− were each presented for 15 periods in a session. The order was randomized but excluded the presentation of the same condition for more than three consecutive periods.

Reinforcement was available at the left key for half of the S+ periods and at the right key for the other half.

The criterion for discrimination was reached when a block of 10 consecutive 1-min. periods was completed under the following conditions: (a) The block contained five positive and five negative periods. (b) At least two of the negative periods were consecutive. (c) S gave no responses to S−. (d) S gave 10 or more responses during each presentation of S+. All except 1 S reached this criterion within 10 sessions. A second generalization test (see below) was administered to this S under the considerations that on Training Sessions 7 through 10, less than 2% of its total responses were to S−, with no responses to S− on as many as four successive S− periods.

In the *simultaneous* discrimination training of Exp. 4, one key was illuminated by S+ and the other by S− during each period. The side was randomized, with the exclusion of more than three successive periods with either arrangement. The criterion for discrimination was reached when a block of five consecutive 1-min. periods was completed under the following conditions: (a)

S− appeared on one key for two periods and on the other key for three periods. (b) S gave no responses to S−. (c) S gave at least 10 responses each period to S+. Of the 8 Ss in Exp. 4, 6 reached criterion within five sessions or less. The other 2 Ss stabilized at about 20% responding to S−. They apparently developed a chained response of pecking at S− and then S+ in rapid succession. This could be reinforced, since no delay of reinforcement was contingent on responding to S−. These 2 Ss were therefore discarded.

*Generalization testing.*—When S reached the criterion for discrimination, the training session was discontinued. On the 2 following days, S was given a generalization test identical to the one it received after VI training in Exp. 1.

## Results

A full description of the course of discrimination training in these experiments has been presented elsewhere (Honig, 1958). Suffice it to say that it differed for the two studies
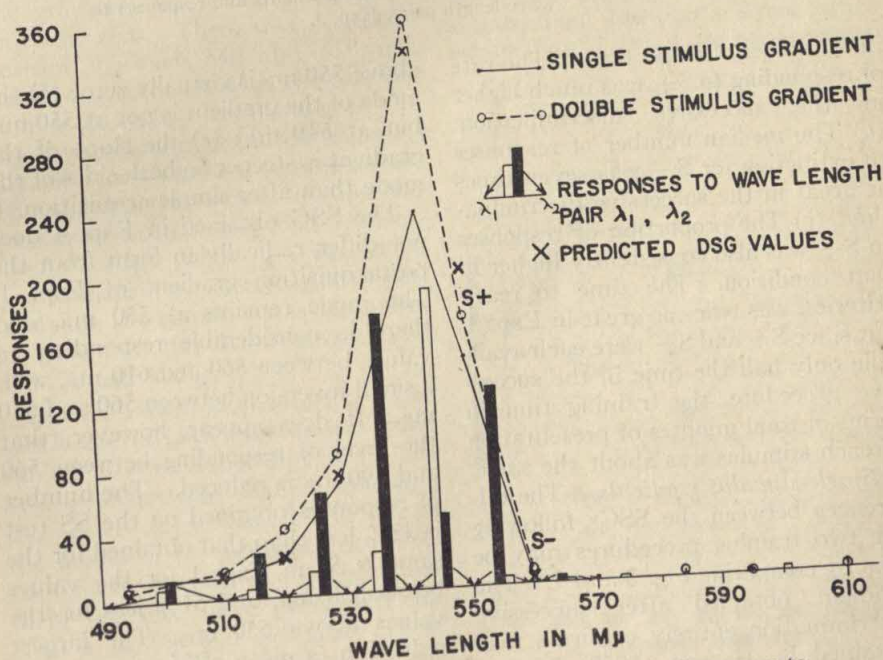


FIG. 3. Single-stimulus and double-stimulus gradients and responses to wave-length pairs, Exp. 3.
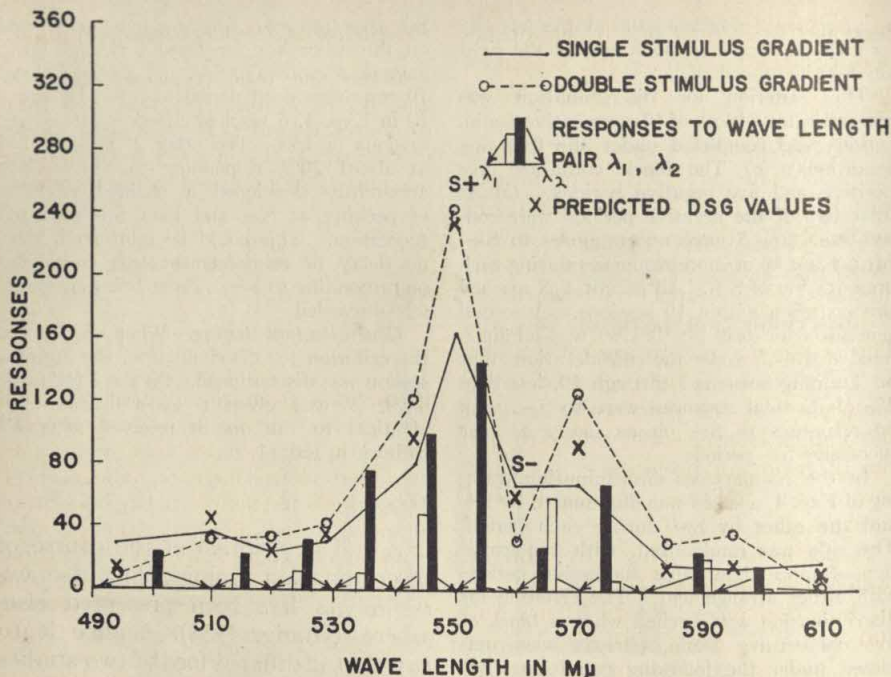
FIG. 4. Single-stimulus and double-stimulus gradients and responses to
wave-length pairs, Exp. 4.

in a number of respects: (a) The rate
of responding to S+ was much higher
in the successive discrimination.
(b) The median number of responses
to extinction for S— was seven times
as great in the successive discrimina-
tion. (c) The proportion of responses
to S— was also consistently higher in
that condition. The time to reach
criterion was twice as great in Exp. 3,
but since S+ and S— were each avail-
able only half the time in the succes-
sive procedure, the training time in
terms of total minutes of presentation
of each stimulus was about the same.

*Single-stimulus gradients.*—The dif-
ferences between the SSGs following
the two training procedures may be
seen by comparing Fig. 3 and 4. The
gradient obtained after successive
discrimination entirely confirms that
obtained by Hanson (1959) in that
(a) the response level at wave lengths

above 550 mμ is virtually zero; (b) the
mode of the gradient is not at 550 mμ,
but at 540 mμ; (c) the slope of the
gradient is steeper on both sides of the
mode than after simple acquisition.

The SSG obtained in Exp. 4 does
not differ radically in form from the
postacquisition gradient of Exp. 1.
The mode remains at 550 mμ, and
there is considerable responding to
values between 560 and 610 mμ, with
a small inversion between 560 and 570
mμ. It does appear, however, that
the level of responding between 560
and 590 mμ is reduced. The number
of responses obtained on the SS test
is 32% less than that obtained for the
same 6 Ss in Exp. 1 for the values
below 550 mμ, and 57% less for the
values above 550 mμ. The largest
reduction (a mean of 75 responses, or
54%) was at 560 mμ, and the reduc-
tions (in terms of absolute amount)

decreased with only one inversion between 560 and 610 m$\mu$. All $S$s showed more reduction at the values above 550 m$\mu$ than below.

*Responding within wave-length pairs: Exp. 3.*—The direction of preference for all the stimulus values conforms to the SSG. This is especially significant for the pair 540, 550 m$\mu$, as the direction of preference is toward the postdiscrimination peak of 540 m$\mu$. The preference of 540 m$\mu$ over S+ is an instance of *transposition of discrimination*. For the pairs comprised of the values between 490 and 540 m$\mu$, the direction of preference is clearly toward S+. These are instances of *transposition-reversal*.

*Experiment 4.*—The direction of preference conforms to the SSG in all but two pairs: 510, 520 m$\mu$ and 590, 610 m$\mu$. Both of these pairs are comprised of values near the ends of the gradient, where the slope is essentially zero. There was no transposition between 540 and 550 m$\mu$, quite in accordance with the absence of a peak shift. Responding to the negative stimulus presented alone on the SS test was not extinguished. The simultaneous discrimination did not transfer completely to the successive situation, while the successive discrimination transferred almost perfectly to the simultaneous situation, as shown by the bars for the 550, 560 m$\mu$ pair in Fig. 3.

*Double-stimulus gradients.*—The DSGs are in both experiments quite similar to the SSGs though on the average somewhat higher, following the pattern set in Exp. 1 and 2. At the modes of the gradients the DSG points are considerably higher, and there are also large differences for Exp. 4 at 560 and 570 m$\mu$. These differences can be understood on the basis of the analysis of DS rates to be presented in the next section.

## Prediction of Double-Stimulus Values

Given the total responses $R_1$ to $\lambda_1$ and $R_2$ to $\lambda_2$ on a SS test, the total responses $r_1$ and $r_2$ for the same values in the pair $\lambda_1$, $\lambda_2$ on a DS test can quite adequately be predicted by the formulae:

$$r_1 = R_1\left(\frac{R_1}{R_1 + R_2}\right)$$

and

$$r_2 = R_2\left(\frac{R_2}{R_1 + R_2}\right)$$

Predictions from these functions have been carried out for each $S$ for each DS value in each experiment to obtain predicted values for $r_1$, $r_2$, and $r_1 + r_2$. Mean predicted are plotted against mean obtained values for each stimulus pair from Exp. 1 and 2 in Fig. 5. The three groups of predicted values, $r_1$, $r_2$, and $r_1 + r_2$, are separated along the vertical axis for clarity. The diagonal lines represent perfect prediction (x = y); deviations from this are indicated by the vertical distance between each point and the line. Identical analyses carried out for Exp. 3 and 4 are not presented
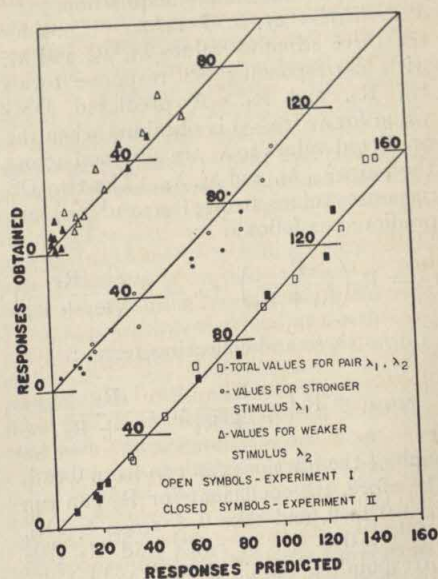


FIG. 5. Predicted and obtained double-stimulus values, Exp. 1 and 2.

here as the general outcome was very similar to that presented for Exp. 1 and 2.

From Fig. 5 it is evident that this method leads to satisfactory results. A goodness of fit test suggested by Lindquist (1953, pp. 344 ff.) was used to test the hypothesis that the obtained do not differ significantly from the predicted values. The error was divided into two components, "vertical placement," or differences between predicted and obtained means, and "departure from pattern," or the error remaining after these means have been equated. Each component provides a mean square which can be divided by the error variance of the obtained data to obtain an $F$ ratio when the data have been corrected for systematic effects, in this case wave length and $Ss$. This analysis was carried out for the sets of $r_1$ and $r_2$ values for the four experiments. The only errors reaching significance are for $r_2$ in Exp. 1, where both vertical placement and departure from pattern are significant at the .001 level. This one obtained significant departure from prediction indicates that by the present method, $r_2$ is systematically underestimated following simple acquisition.

*Prediction of DSG values.*—Consider the three stimulus values $\lambda_1$, $\lambda_2$, and $\lambda_3$, with corresponding SS response totals $R_1$, $R_2$, and $R_3$. A predicted DSG value for $\lambda_2$ ($r_{2DSG}$) is obtained when the predicted values to $\lambda_2$ are summed across the pairs $\lambda_1$, $\lambda_2$ and $\lambda_2$, $\lambda_3$. The two DS response values to $\lambda_2$ ($r_{2a}$ and $r_{2b}$) are predicted as follows:

$$r_{2a} = R_2\left(\frac{R_2}{R_1 + R_2}\right) \quad r_{2b} = R_2\left(\frac{R_2}{R_2 + R_3}\right)$$

Adding these and collecting terms:

$$r_{2DSG} = R_2\left(\frac{R_2}{R_1 + R_2} + \frac{R_2}{R_2 + R_3}\right)$$

Each of the fractions can run from 0 to 1. Therefore, the coefficient for $R_2$ can run from 0 to 2, and thus $0 \leqslant r_{2DSG} \leqslant 2R_2$. The relative size of $r_{2DSG}$ and $R_2$ (the corresponding obtained SSG value) depends on the sum of the two fractions, and this in turn depends on the shape

of the gradient at $\lambda_1$, $\lambda_2$, and $\lambda_3$. In the case, for example, where $\lambda_2$ is a peak, each of the fractions must be greater than .5 since $R_2$ is greater than $R_1$ and $R_3$; the predicted DSG value must therefore be higher than the SSG value. In general, the two fractions will add to 1 (thus providing equal SSG and predicted DSG values) only if the points on the gradient form a geometric series.[3] If the gradient is rising in linear fashion (arithmetic series) the predicted DSG points will lie above the SSG values. The predicted DSG values are indicated as crosses in Fig. 1, 3, and 4 for Exp. 1, 3, and 4. A comparison with the obtained values supports the present analysis; particularly where differences between the DSGs and SSGs are predicted, they are also obtained. The DSG peaks, both predicted and obtained, are considerably higher than the SSG peaks, and for the central values, where the gradients tend to be linear rather than geometric, the DSG lies above the SSG.

## DISCUSSION

The foregoing analysis indicates that the direction and degree of preference between stimuli lying on the generalization gradient can be predicted from the independent response strengths of the stimuli. The occurrence of transposition in conjunction with a shift in the mode of the gradient follows from the general predictive principles outlined above. In this respect, Spence's (1937) analysis of transposition and his prediction of transposition-reversal have been supported. On the other hand, the difference in results between the successive discrimination, which produced a peak shift and transposition, and the simultaneous, which did not, are not anticipated by Spence. While equivalent criteria were demanded at the end of discrimination training under both procedures, in that $S$ did not respond to $S-$ for five successive periods, it appears from Fig. 4 that simultaneous discrimi-

[3] This can be demonstrated algebraically by a proof not given here.

nation training did not result in extinction to S— presented alone. In this discrimination, it was only necessary for S— to become a cue for switching to S+ for perfect discrimination to be manifested. In the successive discrimination, of course, S had no alternative response to S— available during the negative periods to obtain reinforcements, which resulted in extinction under that stimulus condition.

It appears that a shift in the peak of the PDG and the concomitant transposition depended in the present study on the development of genuine extinction to S— alone, rather than a preference for S+. But the critical factor responsible for this extinction need not have been in the manner of presentation of the discriminanda; there is nothing about the simultaneous presentation of S+ and S— per se that necessarily prevented transposition. To evaluate the factor of the manner of stimulus presentation, it would be necessary to develop a procedure where all other aspects of the simultaneous and successive problems are kept equivalent, such as the reinforcement schedules, the alternatives available for responding to S—, and so forth. Thompson (1955) and Riley, Ring, and Thomas (1960) have in effect done this by making both stimuli available on each learning trial, but preventing the animal from comparing the appropriate discriminanda in the "successive" case. They obtained more transposition when a comparison was possible. Whether the same would hold true for the pigeon working with the spectral dimension is an open question.

While the arithmetic model presented above for the prediction of DS from SS values is satisfactory for that purpose, it is not derived from assumptions about, or observations on, specific behavior patterns in the choice situations. It is possible, however, to identify the terms entering the predictive formulae with basic response dimensions in the choice situation, and thus to suggest a pattern of behavior that can be verified. Assume that in the choice situation the pigeon responds to each stimulus value at the same *rate* as when that stimulus is presented alone. This would result in a DS response total of $R_1$ to $\lambda_1$ if there were no interference from the presence of the other choice stimulus. Assume further that the other stimulus does interfere in that its presence reduces the total *duration* of responding to $\lambda_1$, and that the total duration of responding to both values is divided in a proportion reflecting the rates to each value. The proportion for $\lambda_1$ would be $\dfrac{R_1}{R_1 + R_2}$, which, when multiplied by $R_1$, provides the formula used above for the prediction of response totals to $\lambda_1$. The corresponding expression for $\lambda_2$ can be similarly derived. This analysis rests, of course, on a clear distinction between operant rate and the duration of the application of a given rate; while this distinction has been supported empirically by Gilbert (1958), it awaits direct confirmation in the present circumstances.

## SUMMARY

The relationship between stimulus preference and the response strength of singly presented stimuli was investigated with the use of the generalization gradient to provide stimuli of different strengths. After being trained to peck at a 550-m$\mu$ stimulus on a VI schedule, pigeons were given two concurrent generalization tests: one consisting of single stimulus values ranging from 490 to 610 m$\mu$, and one consisting of pairs of such values. The direction of preference within each pair was found to be in direct accordance with the number of responses obtained on the single values during the generalization test. Two groups of birds then received discrimination training between 550 m$\mu$ as S+ and 560 m$\mu$ as S—, one with successive stimulus presentations and one with simultaneous. After this, the generalization and preference tests were administered a second time. The successive discrimination produced a shift in the mode of the gradient away from S— and a concurrent transposition of the discrimination between S+ and the new mode, with transposition-reversal beyond that mode. The simultaneous discrimination produced little change in the gradient and no concurrent transposition.

An arithmetic model is proposed for the prediction of responses to members of stimulus pairs from single-stimulus values. The determination of stimulus preference by the generalization gradient follows directly, and transposition is seen to be no more than a special case. The differential results obtained from successive and simultaneous discriminations are discussed with reference to the actual extinction to S− obtained from each procedure. The operant situation is analyzed in terms of some simple response dimensions to provide a basis for the arithmetic model from which double-stimulus values are predicted.

## REFERENCES

BAKER, R. A., & LAWRENCE, D. H. The differential effects of simultaneous and successive stimulus presentation on transposition. *J. comp. physiol. Psychol.*, 1951, 44, 378–382.

BLOUGH, D. S. Spectral sensitivity in the pigeon. *J. Opt. Soc. Amer.*, 1957, 47, 827–833.

EHRENFREUND, D. A study of the transposition gradient. *J. exp. Psychol.*, 1952, 43, 81–87.

GILBERT, T. F. Fundamental dimensional properties of the operant. *Psychol. Rev.*, 1958, 65, 272–282.

GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, 42, 365–375.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, 51, 79–88.

HANSON, H. M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, 58, 321–334.

HONIG, W. K. Prediction of preference, transposition, and transposition-reversal from the generalization gradient. Unpublished doctoral dissertation, Duke University, 1958.

HONIG, W. K., THOMAS, D. R., & GUTTMAN, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *J. exp. Psychol.*, 1959, 58, 145–152.

KENDLER, T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, 40, 552–562.

LINDQUIST, M. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

RILEY, D. A., RING, K., & THOMAS, J. The effect of stimulus comparison on discrimination learning and transposition. *J. comp. physiol. Psychol.*, 1960, 54, 415–421.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430–444.

THOMPSON, R. Transposition in the white rat as a function of stimulus comparison. *J. exp. Psychol.*, 1955, 50, 185–190.

# RESISTANCE TO EXTINCTION AS A FUNCTION OF AGE AND SCHEDULES OF REINFORCEMENT [1]

NORMAN KASS [2]

*University of Minnesota*

The purpose of this study is to investigate developmental changes in resistance to extinction as a function of the percentage of reinforcement provided during acquisition. Relatively little information is available concerning the effects of partial reinforcement on extinction with children, and no studies investigating age changes have been reported.

There is reason to assume that the development of cognitive processes and the growth of experience with increasing age would result in increasingly rapid recognition of the change in reinforcement schedule between an acquisition and extinction period in a simple operant task, with a resulting decrease in the resistance to extinction following partial reinforcement. Subjects at four age levels were tested, and in order to increase the generality of the findings, four partial reinforcement schedules were employed. Additional groups received 0% and 100% reinforcement.

## METHOD

*Subjects.*—The Ss were 216 preschool and elementary school boys and girls, approximately evenly divided by sex. The Ss were obtained from the Institute of Child Development Nursery School, the Village Nursery

School, and the Elementary School of the University of Minnesota; and from a public elementary school in Minneapolis. All schools enrolled Ss of approximately the same intellectual and socioeconomic levels. All Ss of the relevant age levels in attendance at the time of testing were used in each school.

*Apparatus.*—The apparatus was a simulated slot machine, consisting of a blue wooden cabinet with dimensions of 12 × 14 × 12 in. Centered on the front of the cabinet and 3½ in. from the top was a circular opening 1 in. in diameter. A large number of pennies was visible behind a piece of clear plastic covering the opening. The box containing the pennies was illuminated with a 7-w. lamp. Pennies were dispensed through a rectangular opening located on the lower front of the cabinet. A ¾-in. wooden shaft, to which a 3 × ½ in. aluminum handle was attached, projected from the front of the cabinet 3 in. from the right and 6¼ in. from the bottom.

The rear of the cabinet was open to allow E to operate the apparatus. A 12 × 14 in. board was attached to each side of the machine to prohibit S from observing E load the machine.

Attached to the inside base of the cabinet was a dispensing device, which consisted of a wheel 11 in. in diameter, with 24 slots circling the outside edge. The wheel was attached to the base by a pivot with a spring-loaded brake which allowed the wheel to turn without spinning freely. This wheel was rotated one slot at a time by means of a pawl. The pawl rode in a track and was driven by a lever attached to the wooden shaft described above. When S pulled the handle, it rotated the wheel one slot forward and also activated a counter to record the response. On the base of the cabinet was an opening over which the wheel rotated with dimensions corresponding to those of the bottom of each slot. This opening was connected by means of a chute to the opening in the front of the machine. The pennies placed in the various slots could be dispensed when a particular slot was in line with the opening in the base. With this arrangement any variable-ratio schedule could be presented. Each machine had to be reloaded after it was played 24 times.

The apparatus was extremely simple to operate, and involved pulling the handle down and releasing it. When the handle was pulled down, the light which illuminated the pennies went out and relit when the handle was returned to its original position.

At the completion of the experiment Ss traded their pennies for prizes, such as toy wrist watches, badges, umbrellas, flutes, and kaleidoscopes.

*Design.*—The variables employed were percentage of reinforcement and chronological age of S. The six reinforcement schedules provided reinforcement on 0%, 16⅔%, 33⅓%, 60%, 80%, and 100% of the acquisition trials. The Ss were selected from four age levels: 4, 6, 8, and 11 yr.

Each S was given an acquisition and an extinction series. The acquisition series consisted of 30 trials. During this period S received, according to his reinforcement schedule, either 0, 5, 10, 18, 24, or 30 reinforcements. The reinforcements were presented randomly with the restriction that all Ss except those in the 0% condition were reinforced on the last trial of the acquisition series.

The study thus employed a 4 × 6 design with 9 Ss in a cell. The Ss at each age level were assigned to the reinforcement conditions at random.

*Procedure.*—The Ss were obtained by E from the classroom and taken to an experimental room located in a quiet section of the building. The S was told that he was going to play the "penny machine game" and was seated before the apparatus. The pennies in the illuminated window were pointed out by E.
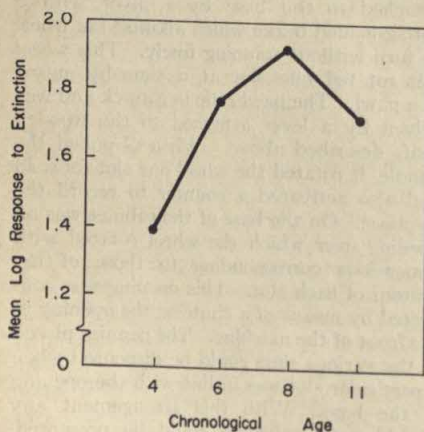
The E said:

You see this machine is full of pennies. Do you see all these pennies? Now in this game you try to get as many pennies out of the machine as you can, because the more pennies you get out of the machine, the better the prize you can buy with the pennies.

The E showed S the prizes he could buy. The E continued:

See all the prizes we have that you can buy with the pennies you get out of the machine. Now I'll show you how to play the game. As long as the light is on, you can pull the handle and the pennies come out of here. You can play the game as long as you want. You don't have to hurry. Just tell me when you want to stop. Remember, the more pennies you get, the better the prize you can buy with your pennies.

The last statement was repeated on Trials 10 and 25 of the acquisition series. There was no indication from E when the acquisition series ended and the extinction series began. Comments and questions from S were ignored by E, and if S persisted, E replied, "You can play the game as long as you like. Tell me when you want to stop playing."

Each S played until he indicated he wanted to stop or after 370 extinction trials, when S was stopped by E. Only 8 Ss failed to complete the 30 acquisition trials and were replaced. All of these Ss were at the 4-yr. age level and were in either the 0% or 100% group.

At the end of the session S was allowed to choose a prize to buy with his pennies. The Ss in the 0% group were told that even though they had not won any pennies, they could have a prize for playing so well. The Ss were asked not to tell other children about the game.

## RESULTS AND DISCUSSION

The primary score used in the analysis of the results was the number of responses made during the extinction period. Because of the heterogeneity of variance among the cells, a log transformation of the scores was performed.

The results related to age differences are shown in Fig. 1, which presents the mean log of the number of responses made during the extinc-



FIG. 1.  Mean log responses to extinction as a function of CA of Ss.

tion period at each CA level for all percentages of reinforcement combined.

The results related to percentage of reinforcement are shown in Fig. 2, which presents the mean log responses to extinction for each percentage of reinforcement for all age levels combined. The results are in line with those obtained with college $S$s by Lewis and Duncan (1956, 1957). There was a consistent decrease in number of responses to extinction with increasing percentage of reinforcement from $16\frac{2}{3}\%$ to $100\%$. The $S$s in the $0\%$ group made somewhat fewer responses during extinction than $S$s in the $16\frac{2}{3}\%$ group, but made a greater number of responses than $S$s in any other group.

In analysis of variance of the transformed scores, differences associated with CA were highly significant ($F = 9.54$, $df = 3/192$, $P < .001$), as were the differences associated with percentage of reinforcement ($F = 15.65$, $df = 5/192$, $P < .001$). The interaction term was not significant ($F = .45$).

The Duncan multiple range test (Duncan, 1955) was used to determine which of the age levels and percentages of reinforcement differed significantly. The scores of $S$s at CA 4 differ significantly ($P < .01$) from those of $S$s at the other CA levels; the scores of the remaining groups do not differ significantly from each other. The following differences according to percentage of reinforcement are significant at the .01 level: $16\frac{2}{3}\%$ vs. each of the larger percentages; $0\%$ vs. $60\%$, $80\%$, $100\%$; $33.3\%$ vs. $80\%$ and $100\%$.

The results may be interpreted by the hypothesis that the rate of extinction is decreased as the discrimination between the acquisition and extinction series becomes more difficult. The maximum similarity between acquisition
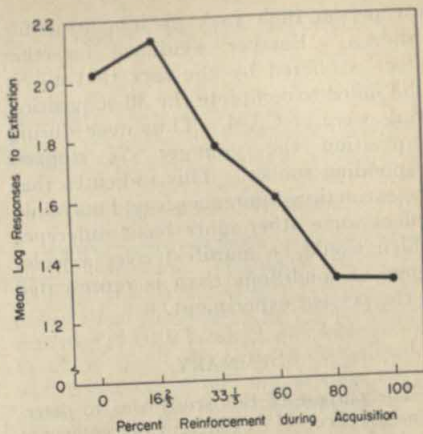


FIG. 2. Mean log responses to extinction as a function of percentage of reinforcement during the acquisition series.

and extinction trials occurs in the $0\%$ group, where there is no way for $S$s to discriminate the end of the first series and the beginning of the second. The maximum dissimilarity occurs in the $100\%$ group. The slight but nonsignificant increase from the $0\%$ to the $16\frac{2}{3}\%$ group provides the only deviation from a consecutively decreasing number of trials to extinction with increasing percentage of reinforcement. The lack of a significant interaction between percentage of reinforcement and CA indicates that, in general, the trend of the results is similar at each age level. There was a curvilinear relationship found between CA and number of responses to extinction. These differences do not agree with those which would be expected from the hypothesis that increasing age results in an increasing ability to discriminate between changes in patterns of reinforcement. According to such an hypothesis the older $S$s would be expected to show the least resistance to extinction. Since the only significant differences were those found between $S$s at CA 4 and $S$s at higher CA levels, the main problem is to determine why the youngest $S$s should extinguish more readily. This finding might be explained in terms of the developmental changes in length of attention span. Preschool children may

not persist in a task as long as older children. Further evidence for this effect is offered by the fact that all Ss who failed to complete the 30 acquisition trials were at CA 4. Thus even during acquisition the younger Ss stopped responding sooner. This indicates that the extinction differences found may only reflect some other more basic difference which would be manifest over a wider range of conditions than is represented in the present experiment.

## SUMMARY

The purpose of this study was to determine the effects of six different percentages of reinforcement upon extinction of a lever-pulling response. Four age levels were employed to determine developmental changes in resistance to extinction. The apparatus employed was a simulated slot machine designed for use with young children.

A total of 216 children at CA 4, 6, 8, and 11 received reinforcement on either 0, $\frac{1}{6}$, $\frac{1}{3}$, $\frac{3}{5}$, $\frac{4}{5}$, or all of the trials of a 30-trial acquisition period. The extinction trials were continued until S wished to stop or until 370 extinction responses had been made. The response measure was the total number of responses to extinction.

A decrease in number of responses during extinction was found with increasing percentages of reinforcement. The least number of trials to extinction was shown by Ss at CA 4 and the greatest by Ss at CA 8. There was no significant interaction between CA and percentage of reinforcement.

## REFERENCES

DUNCAN, D. B. Multiple range test and multiple F tests. *Biometrics*, 1955, 11, 1–42.

LEWIS, D. J., & DUNCAN, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *J. exp. Psychol.*, 1956, 52, 23–27.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as functions of percentage of reinforcement and amount of reward. *J. exp. Psychol.*, 1957, 54, 115–120.

# SUBJECTIVE SCALE OF FORCE FOR A LARGE MUSCLE GROUP [1]

## HANNES EISLER [2]

*Harvard University*

Although subjective force is one of the classical continua studied in psychophysics, most studies of force have been limited to lifted weights. Recently, however, Stevens and Mack (1959) scaled subjective force, as exerted in the squeezing of a handle. Their experiments confirmed the psychophysical power law; the exponent obtained was 1.7. Borg and Dahlström (1960) investigated muscular work carried out on a bicycle ergometer. Though the variable they studied was power rather than force, their experiments can also be included in investigations of muscular effort. They found an exponent of 1.6.

The investigation reported here consists of a series of experiments carried out to scale the subjective force exerted by a comparatively large muscle group. The muscle group chosen contains those muscles that are used in pushing or pressing a pedal with the foot in a horizontal, forward direction. In addition to being intrinsically interesting, these experiments make it possible to compare the subjective scale of force exerted by the large leg muscles with the scale obtained for the smaller muscles of the forearm used in squeezing a hand dynamometer. To summarize the results, a power function was obtained for foot pressure also, and its exponent did not differ appreciably from the exponent for handgrip.

## GENERAL METHOD

*Scaling.*—Five methods were used: magnitude estimation, magnitude production, matching of force of handgrip to force of foot pressure and the reverse, and cross-modality matching of both handgrip and foot pressure to white noise. The magnitude estimation was repeated once, and the matching of foot pressure to handgrip twice with slight variations, so that, all in all, eight experiments enter into the present study. Except in the first magnitude estimation experiment in which $O$ gave four judgments for every stimulus, two values for every stimulus were obtained from each $O$ in all experiments.

*Subjects.*—Twelve $O$s were used in each experiment. The groups largely overlapped and were identical for the first magnitude estimations and the magnitude productions, as well as for all the experiments in which foot pressure and handgrip were matched. Altogether, 16 men and 5 women took part in the investigation, mostly graduate students of psychology and staff members.

*Apparatus.*—In all the experiments, $O$ sat in a rigid chair and with his right foot pressed a pedal in a forward, horizontal direction. The distance through which the pedal moved was very small.

In the experiments with magnitude estimation and the matching of handgrip to foot pressure, the pedal was connected through a lever system to the platform of a large beam scale. The measuring beam of the scale was loaded with a given *stimulus* weight, and $O$, by pressing the pedal, brought the scale into equilibrium. Equilibrium was indicated to him by a pointer on a display.

In the experiment in which force of foot pressure was matched to force of handgrip, a handle was connected with the platform of the scale. Forces exerted as *responses* (foot pressure in magnitude production, foot pressure when matched to handgrip, etc.) were measured with tensile gauges connected with the pedal, the handle, or both, depending on the particular experiment.

When force of foot pressure and force of handgrip were matched to white noise, $O$s listened to the band of noise (75 to 2400 cps) through a pair of earphones.

*Sources of error.*—Three sources of error were evident in the series of experiments reported, though not every error was to be found in every experiment:

1. The pointer in the display could not be kept completely stationary. There seemed to be some correlation between the speed of its movement and the force exerted. This movement may or may not have influenced *O*'s judgment. Since this source of error did not exist in all the experiments and since the agreement among all the experiments is quite good, it is probably of only minor importance.

2. People differ in strength. The problem caused by this state of affairs could be attacked by (*a*) limiting the range investigated to the maximum force of the weakest *O*, but restriction of range is not an appealing solution; (*b*) picking a particular sample of strong *O*s, but restriction of sample is not an attractive solution either; (*c*) or dropping *O*s as the forces required were increased. This solution was chosen. It is a legitimate solution if the exponent of the power function is independent of the strength of *O*. In that case the only effect would be a decrease in the reliability for strong stimuli, since the points obtained are based on fewer *O*s.

3. While he pushed the pedal, *O* had to hold his leg up by his own force. For small stimuli the force required was apparently greater than the force of pushing. For large stimuli the friction between the sole of the shoe and the pedal appeared to be sufficient to eliminate this extra force. The *O*s seemed to be trying to neglect the holding force and only to take the pushing force into account.

## MAGNITUDE ESTIMATION AND MAGNITUDE PRODUCTION

### Method

Whether the function obtained by magnitude estimation was independent of the strength of *O*s was tested crudely in two ways. (*a*) The *O*s were divided into a group of 7 who were able to exert the greatest force used, 400 lb., and a second group of 5 who were not. Medians of the magnitude estimates of the six lowest stimuli common to the two groups were computed for each group separately. A plot of the medians of the two groups against each other was essentially linear. (*b*) Medians were computed, for each group separately, of the magnitude estimates for the highest stimulus force each *O* could exert, the second highest stimulus presented, and so on. This procedure is meaningful when the stimuli are spaced logarithmically, on the assumption that the power law holds for this case. Again, when the values for one group are plotted against those for the other, the result is essentially linear. Thus the function relating subjective force to physical force appears to be largely independent of the strength of the particular *O*.

The fact that a plot of the logarithms of subjective force (geometric means) against the logarithms of force in pounds was somewhat curved indicates that the "threshold" cannot be neglected. Consequently, the values were corrected for "effective threshold" or "subjective zero," by a method developed by Ekman (1961).

### Results

The data of the magnitude estimation experiment indicated that subjective force is a power function by Ekman's criterion. In contrast to most other scaling experiments, however, this one gave a *negative* threshold. The threshold correction amounted to $+8.9$ lb. The logarithms of the magnitude estimates are plotted in Fig. 1 against the logarithms of the corrected (circles) and uncorrected (crosses) physical values. A least squares fit yielded a slope, that is to say, an exponent of 1.51.

The magnitude production data were treated in the same way as the magnitude estimation data. For mag-
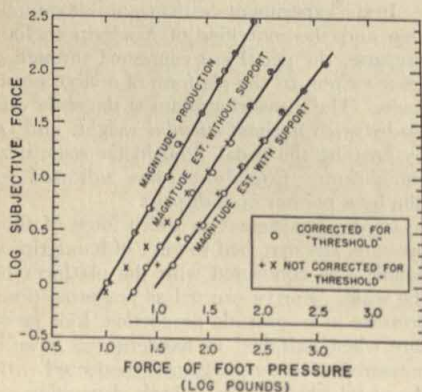


FIG. 1. Subjective force of foot pressure in logarithmic units as a function of physical force in log pounds. (The upper scale on the abscissa refers to the magnitude estimation experiment in which *O*'s leg was supported.)

nitude production, the threshold was positive and the correction small, −1.8 lb. The data are also shown in Fig. 1. The exponent was 1.70.

It is common to find this discrepancy between the exponents for magnitude estimation and magnitude production; S. S. Stevens has called it the "regression effect" (see e.g., Reynolds & Stevens, 1960). The O decreases the range of the variable over which he has control—numbers in magnitude estimation and forces in magnitude production.

The consensus of these two experiments is that, after application of threshold corrections, subjective force of foot pressure grows as a power function of physical force with an exponent of approximately 1.6. The result agrees with the findings of Borg and Dahlström (1960), derived from fractionation experiments. They too obtained the exponent 1.6 (and a negative threshold).

In order to find out whether the negative threshold found in the magnitude estimation experiment was related to the third source of error mentioned above—the force required to hold up the leg—the magnitude estimation experiment was repeated with the variation that O's heel was supported by a leather belt. The results are given in Fig. 1. The negative threshold was confirmed, although the threshold correction was smaller in this experiment, +3.8 lb. The exponent, however, decreased to 1.31. Since Os were divided about whether the supporting belt was an improvement, and some of them found it more uncomfortable than the first experiment, the results of this experiment will not be taken into account in the conclusion of the whole series of experiments.

## MATCHING EXPERIMENTS

### Method

The procedure used in the experiment in which handgrip was matched to foot pressure was similar to that of the experiment with magnitude estimation, except that, instead of responding with numbers, O now equated squeezes on a handle with the force exerted by pressing the pedal.

Matching foot pressure to handgrip is the inverse of the procedure described above. Three variations were carried out. Two of them differed from each other only in the stimuli chosen and the leverage applied between handle and scale. The leverage was changed in such a way as to increase the precision of the forces presented by E. In the third variation the Os were asked to press the pedal, not simultaneously with the hand squeezes, but after them.

In order that the results for simultaneous and successive matching could be compared, the medians of the two experiments were plotted against each other. Since a straight line was obtained, indicating that successive matching made no difference, the results from the experiment with successive matching were not treated further.

In the final experiment, Os were asked to match in random succession both force of foot pressure and force of handgrip to a band of white noise, whose intensity varied between 30 and 100 db. re 0.0002 dyne/cm². The point of this experiment was to match force of foot pressure with force of handgrip indirectly. A noise of a certain intensity was presented and O was informed whether he was to press the pedal or to squeeze the handle. This was the only experiment in which the weakest Os did not drop out as the forces were increased.

### Results

All the matching experiments gave curvilinear functions when force of handgrip was plotted against force of foot pressure in log-log coordinates (geometric means). (For the experiment with white noise, force of handgrip matched to a certain noise intensity was plotted against the force of foot pressure matched to the same noise intensity.) Plotted linearly, however, straight lines (that did not pass through the origin) were obtained. This outcome indicated (a) a nonzero threshold for at least one of the continua (with no way to tease out the separate threshold corrections), and (b) a power relation between subjective force of foot pressure and subjective force of handgrip, with an exponent of 1. Thus, if either of the

two continua grows as a power function of physical magnitude, the other grows with the same exponent.

The straight lines were computed according to a variation of the method of least squares. The customary method weights the high values more than the low values. Because of the comparatively great ranges covered, the application of that method would have introduced a heavy bias. Therefore the sum of squares of the *relative* deviations,

$$\Sigma[(y-y_1)/y]^2$$

where $y$ equals experimental values and $y_1$ computed ones, has been minimized rather than the sum of squares of the *absolute* deviations,

$$\Sigma(y-y_1)^2.$$

The lines obtained were normalized so that they passed through the origin with a slope of 1, and they are sum-

FIG. 3. Force of handgrip (in log pounds) as a function of force of foot pressure (in log pounds) from four matching experiments. (The values of foot pressure are corrected for "threshold.")

marized in Fig. 2. Figure 3 shows the outcome of the individual matching experiments in log-log plots after the threshold correction has been arbitrarily carried out on the foot pressure continuum. The lines drawn in Fig. 3 have a slope of 1.

It is of interest to notice that the parameter $k$ in Stevens' power law

$$\psi = k\phi^n$$

where $\psi$ refers to a subjective magnitude and $\phi$ to a physical magnitude, seems to depend on the particular experimental procedure. In matching, two sets of subjective magnitudes are equated (Stevens, 1959):

$$\psi_1 = k_1\phi_1^{n_1}$$
$$\psi_2 = k_2\phi_2^{n_2}.$$

In the present case, since $n_h = n_f$, this yields

$$\phi_h = (k_f/k_h)\phi_f$$

where the subscripts $h$ and $f$ refer to handgrip and foot pressure. It is the factor $k_f/k_h$ whose logarithm constitutes the intercepts in Fig. 3. Note that the data of both experiments in which foot pressure was matched to handgrip scatter

FIG. 2. Force of handgrip as a function of force of foot pressure from four matching experiments. (The values are normalized such that the four different straight lines obtained by a variation of the method of least squares are made to coincide as one line, which passes through the origin and has a slope of 1. Because the range covered is great, it is divided into two sections which are superimposed.)

around the same straight line, whereas the other two matching experiments yield different intercepts. Some kind of rule of least effort seems to hold, more conspicuously for foot pressure: the forces over which $O$ has control are small relative to those presented by $E$, thus yielding a large coefficient $k$; when both forces are matched to white noise, the ratio $k_f/k_h$ becomes almost 1.

## SUMMARY

The subjective force of pushing a pedal with the leg has been scaled as an instance of the subjective force exerted by a large muscle group. The following methods were employed: magnitude estimation, magnitude production, matching the force of handgrip to the force of foot pressure and vise versa, and matching both foot pressure and handgrip to the intensity of white noise. The experiments involving numbers yielded a power function with an exponent of 1.6 relating subjective force to physical force. All the matching experiments showed that the exponent for force of foot pressure and force of

handgrip is the same. The exponent for handgrip has previously been determined as 1.7. Thus the subjective force of foot pressure, as measured in this study, approximates a power function of physical force not appreciably different from the one found earlier for force of handgrip. The exponent for foot pressure approximates 1.65.

## REFERENCES

BORG, G., & DAHLSTRÖM, U.  The perception of muscular work.  Publication No. 5, 1960, Umeå Research Library, Sweden.

EKMAN, G.  A simple method for fitting psychophysical power functions.  *J. Psychol.*, 1961, **51**, 343–350.

REYNOLDS, G. S., & STEVENS, S. S.  The binaural summation of loudness.  *J. Acoust. Soc. Amer.*, 1960, **32**, 1337–1344.

STEVENS, J. C., & MACK, J. D.  Scales of apparent force.  *J. exp. Psychol.*, 1959, **58**, 405–413.

STEVENS, S. S.  Cross-modality validation of subjective scales for loudness, vibration, and electric shock.  *J. exp. Psychol.*, 1959, **57**, 201–209.

# RE-EXAMINATION OF THE SERIAL POSITION EFFECT[1]

MURRAY GLANZER

*Department of Psychiatry, University of Maryland*

AND STANLEY C. PETERS [2]

*Walter Reed Army Institute of Research*

Existing generalizations about the serial position effect in rote learning have recently been challenged as based on artifacts. McCrary and Hunter (1953) presented evidence that the different curves produced by experimental variables, such as presentation rate, distribution, and list difficulty, are a single curve multiplied by different numbers of errors. Their contention was borne out by the findings of Braun and Heymann (1958). A re-examination of the serial position effect seems necessary. Since any statement about serial position effects must refer to the beginning and end of a series of items, the first step in the re-examination is to analyze the meaning of the terms *beginning* and *end*. The analysis consists of three steps: to determine what characteristics $E$ points to as defining the beginning and the end of a list, to analyze these characteristics, and to determine their effect on $S$'s performance.

In a rote learning experiment, $E$ presents to $S$ a continuous and repetitive cycle of events: the series of syllables, a gap; the series of syllables, a gap. The terms *beginning* and *end* are used to refer to the characteristics associated with this gap. They are the following: (*a*) Primacy-recency: The first item $S$ sees, and the last item he sees before the cycle repeats

itself, appear on either side of the gap, because $E$s usually present the list starting from the gap. (*b*) Spacing: The appearance of the gap coincides with a period in which $S$ is not required to anticipate syllables. (*c*) Association break: Every item in the series is both a stimulus and a response item, except the two on either side of the gap. The $S$ is not required to form an association across the gap. The first item in a list is usually a pure stimulus item. The last item is a pure response item, in that it does not function as stimulus for the first item.

The purpose of the following experiments was to separate the factors that make up the complex called *the beginning* and *the end* of the list, and to determine the part these factors play in the serial position effect.

## EXPERIMENT I: PRIMACY-RECENCY

Mitchell's (1934) evidence that the usual bowed curve appears in lists that have neither spacing nor association breaks would indicate a considerable role for primacy-recency effects. In Exp. I, the role of primacy-recency was evaluated by varying $S$s' starting positions: starting some $S$s at the normal position in the list, and starting other $S$s at what would ordinarily be called the middle of the list. The normal position in the list is that part which has the gap followed by the asterisk, or other starting cue; this may be called the *structural beginning*. The *temporal*

*beginning*, a term which may be used to refer to the first syllable exposed to *S*, may or may not coincide with the structural beginning. The two experimental conditions here were one in which the structural and temporal beginning coincided, and one in which they did not coincide. If the temporal starting point (the primacy-recency factor) plays an important role in determining the serial position effect, then groups starting at the structural middle should show a flattened or indented serial position curve.

## Method

*Subjects.*—The *S*s were 32 Army medical service enlisted men. They were average or above average in their scores on Army intelligence tests (General Technical score of the Army Classification Battery).

*Materials.*—Two six-syllable lists (Glaze 20% association value) were used: List I: *, ZID, NUK, WEF, QAM, TUH, BEJ; List II: *, QAM, TUH, BEJ, ZID, NUK, WEF. The two lists consist of the same series of syllables, with the structural beginning placed at different positions.

*Procedure.*—The *S*s were first given two three-syllable practice lists (Glaze 100% association value). One practice list was started from the structural beginning, i.e., the asterisk. The other was started from the structural middle. Half the *S*s were started from the structural beginning in their first practice list; the other half were started from the structural middle in their first practice list. After learning both practice lists to a criterion of one perfect trial, *S* learned one of the six-syllable lists.

The presentation rate was 3 sec. per syllable, with a 6-sec. interval (two blank spaces on the drum) between successive cycles of the list. All groups received the same instructions: to try to anticipate each syllable before it appeared. No reference was made to either the beginning or the end of the list. The list was learned to a criterion of three consecutive errorless trials. Half the *S*s learned List I; the other half, List II. Half the *S*s started at the structural beginning with the asterisk exposed. The other half started at the structural middle of the list, with either WEF or BEJ exposed. Trials were counted as starting at the structural middle of the list for the *S*s who started at that point.

## Results

Since the work of McCrary and Hunter (1953) and Braun and Heymann (1958) showed that the total number of errors has a multiplicative effect on the shape of the serial position curve, here, and in the subsequent experiments, error scores were translated into logarithms. This translation converts the multiplicative effect to an additive factor and permits direct comparison of the shapes of the curves. This procedure was adopted instead of dividing through by the total number of errors, as in the work mentioned above.

The groups that started in the structural middle had a slightly higher serial position curve (see Fig. 1) than those starting at the structural beginning. The overall differences between the two experimental groups
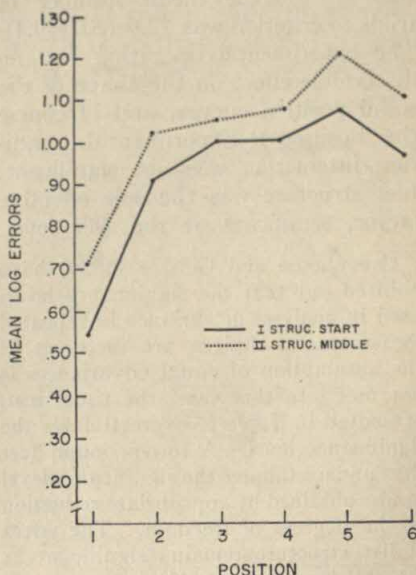


FIG. 1. Serial position curves with primacy-recency varied: Exp. I. (Group II started in the middle of the list.)

TABLE 1

ANALYSIS OF VARIANCE OF LOG ERROR
SCORES: EXP. I

| Source | df | MS | F |
|---|---|---|---|
| Between $Ss$ | | | |
| Condition (C) | 1 | .5645 | 1.69 |
| Lists (L) | 1 | .3934 | |
| C × L | 1 | .1068 | |
| Error (between) | 28 | .3331 | |
| Within $Ss$ | | | |
| Position (P) | 5 | .9834 | 42.76*** |
| P × C | 5 | .0171 | |
| P × L | 5 | .0193 | |
| P × C × L | 5 | .0384 | 1.67 |
| Error (within) | 140 | .0230 | |

*** $P < .001$.

were not, however, statistically significant (see Table 1). For the group starting at the asterisk, the overall mean of the converted error scores was 5.5 ($SD = 1.4$); the mean number of trials to criterion was 19.7 ($SD = 9.2$). For the group starting at the structural middle, the mean of the converted errors was 6.1 ($SD = 1.4$); the mean number of trials to criterion was 22.9 ($SD = 9.4$). The experimental operation had no discernible effect on the shape of the serial position curves, and of course the Position × Experimental Condition interaction was not significant. List structure was the sole effective factor, significant at the .001 level.

Greenhouse and Geisser (1959) have pointed out that the significance levels used in analyses of variance in repeated measurements designs are incorrect, if the assumption of equal covariances is not met. In that case, the usual test, presented in Table 1, overestimates the significance level. A lower bound test that underestimates the significance level can be obtained by appropriate reduction of the degrees of freedom. The effect of list structure remains significant at the .001 level ($df = 1/28$).

The lists used in this experiment were shorter than the 10- to 12-syllable lists usually used in serial position studies.

However, the curves (Fig. 1) show the same skewed bow shape as the curves obtained with longer lists (Hovland, 1938a). They are also similar to the family of curves obtained by Robinson and Brown (1926) for lists that range from 5 to 17 syllables. There is no basis, therefore, for assuming that different factors are at work in 6-syllable lists than those in longer lists.

The type of scoring used above (based on the total number of errors before the criterion was met) may conceal an early effect of primacy-recency that is swamped by the repeated effect of list structure. To check on this possibility, the number of the first trial on which a correct anticipation occurred was scored for each position in the list. With the starting position at the asterisk, the means for the six positions were 2.38, 5.00, 8.06, 8.50, 9.75, and 6.62. With the starting position in the middle of the list, the means were 3.81, 6.38, 6.31, 7.81, 11.31, and 8.50. There is an indication of flattening of the curve for the lists started at the middle. The Position × Condition interaction, however, was not significant ($df = 5/140, P > .10$).

It is clear that the structure of the list, rather than a primacy-recency effect, is the major determinant of the serial position effect; the same serial position effect appears with or without primacy-recency. The results do not contradict Mitchell's (1934) results, since in her lists the structural characteristics of spacing and association break were absent. As will be pointed out below, when the structural characteristics are removed, primacy-recency effects can be seen. For the usual type of list presentation, however, the important factor seems to be the structure of the list. The next step was to examine the two factors associated with the structural beginning and end of the list: spacing and the association break.

EXPERIMENT II: SPACING

If spacing is a factor determining the appearance of serial position effect, then changing the amount of spacing

should affect the shape of the serial position curve. More specifically, if the confounding multiplicative effect of total number of errors is eliminated, then increasing the spacing should give more pronounced serial position effect. McCrary and Hunter (1953) and Braun and Heymann (1958) have presented evidence that contradicts this hypothesis, but they compared only 6-sec. and 126-sec. intertrial intervals. Intervals between 0 and 16 sec. were included in the experiment below. Perhaps more important than the range of spacing values is the fact that in the Hovland (1938b, 1940) studies re-analyzed by McCrary and Hunter, and also in the Braun and Heymann study, the long and short spacing intervals were not completely comparable. In those studies, an interpolated task was given during the 126-sec. interval, but not during the 6-sec. interval.

## Method

*Subjects.*—The Ss were 90 Army medical service enlisted men. The group was above average on Army intelligence tests (General Technical score on the Army Classification Battery).

*Materials.*—Two 10-syllable (Glaze 20% association value) lists were used: List I: *, RUW, GIY, POH, WEF, QAM, ZIX, NUK, BEJ, XOC, KAZ; List II: *, ZIX, NUK, BEJ, XOC, KAZ, RUW, GIY, POH, WEF, QAM.

*Procedure.*—The Ss were assigned in equal numbers to each of the two list conditions and the five spacing conditions. They first learned two 3-syllable practice lists (Glaze 100% value) to a criterion of one perfect trial, under the same spacing condition as their main list. They then learned one of the 10-syllable lists to a criterion of two errorless trials. The criterion was set as high as possible within the time available, so that the serial position curves would represent final, stabilized performance.

The lists were presented at a 2-sec. rate, with zero, one, two, four, or eight blank spaces (0, 2, 4, 8, or 16 sec.) between the end of the list and the reappearance of the beginning. There were 9 Ss who were unable to

reach the criterion within the 75 min. available: 2 each in the zero- and one-space condition, 1 each in the two- and eight-space condition, and 3 in the four-space condition. These Ss were replaced by others.

## Results

The mean log error curves show an orderly change from a relatively flat curve for the zero-spacing condition, to a sharply peaked curve for the maximum spacing condition (see Fig. 2). The statistical evaluation (see Table 2) of the effect with the usual tests, however, does not find the Spacing × Position interaction significant (.10 > P > .05). This analysis does not, of course, reflect the orderly nature of the changes in the curves. For example, there is a progressive increase in the slope of the curve between Positions 1 and 6 as the spacing increases. The rank order correlations for the difference between Positions 1 and 6 with the amount of spacing is 1.00 for both the List I and List II groups. In both groups, there is a corresponding, but less marked, decrease in slope between
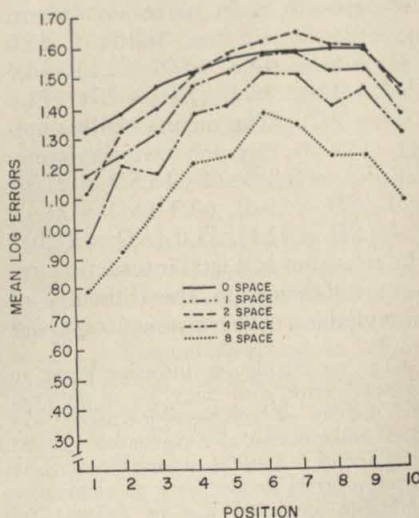


FIG. 2. Serial position curves with spacing varied: Exp. II.

### TABLE 2

ANALYSIS OF VARIANCE OF LOG ERROR
SCORES: EXP. II

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | | | |
|   Spacing (S) | 4 | 3.5962 | 7.00*** |
|   Lists (L) | 1 | 0.0059 | |
|   S × L | 4 | 0.3396 | |
|   Error (between) | 80 | 0.5138 | |
| Within Ss | | | |
|   Position (P) | 9 | 2.0380 | 90.18*** |
|   P × S | 36 | 0.0315 | 1.39 |
|   P × L | 9 | 0.0437 | 1.93* |
|   P × S × L | 36 | 0.0168 | |
|   Error (within) | 720 | 0.0226 | |

\* $P < .05$.
\*\*\* $P < .001$.

Positions 6 and 10 as spacing increases. The rank order correlation for the difference between Positions 6 and 10 with the amount of spacing, is .70 in both groups.

The analysis demonstrates the expected overall differences between the spacing conditions, the means declining ($P < .001$) with increased spacing, except for a reversal between the one- and two-space conditions. The overall means of the converted errors for the zero- to eight-space conditions, in order, were as follows: 15.0 ($SD = 2.1$), 14.3 ($SD = 2.1$), 14.8 ($SD = 1.9$), 13.4 ($SD = 2.7$), 11.6 ($SD = 2.2$). The means of the number of trials to reach criterion were, in order, as follows: 58.4 ($SD = 21.8$), 54.4 ($SD = 19.6$), 60.3 ($SD = 21.4$), 47.7 ($SD = 22.1$), 35.0 ($SD = 18.0$).[3] The Position × Lists interaction reflects differences in the difficulty of the syllables. The absence of a sig-

[3] The contribution of intrusion errors to the total error score may be of interest. The number of complete intrusions was tallied and expressed as a percentage of total errors for each S. The mean percentage of intrusion errors for the zero- to eight-space conditions, in order, was as follows: 5.3 ($SD = 6.1$), 4.7 ($SD = 6.5$), 3.0 ($SD = 2.5$), 4.2 ($SD = 2.7$), 7.7 ($SD = 6.7$).

nificant third-order interaction indicates that the effect of the spacing on the serial position curve does not differ for the two lists. The position effect is significant ($P < .001$) and remains at that level under the Greenhouse-Geisser lower bound test.

The results lent some support to the hypothesis that the serial position curve changes in an orderly fashion as a function of spacing. The statistical support was, however, not satisfactory.

### EXPERIMENT III: SPACING AND ASSOCIATION BREAK

In this experiment, the aim was to retest the effect of spacing on the serial position effect and also to evaluate the effect of the association break. The expectation was that spacing would increase, and that associative chaining across the gap would decrease, the serial position effect.

### Method

*Subjects.*—The Ss were 120 college students. They were paid for their participation.

*Materials.*—In order to insert associative chaining between successive presentations of the list, the lists were constructed in quasi-paired-associates format. Ten chained lists were constructed by rotating the syllables systematically through the 10 available positions. Two of the chained lists are shown below.

Chained 1: \* ZIX NUK \* NUK BEJ \* BEJ XOC \* XOC KAZ \* KAZ RUW \* RUW GIY \* GIY POH \* POH WEF \* WEF QAM \* QAM ZIX (Space 0, 2, or 8) \* ZIX NUK \* (etc.)

Chained 2: \* NUK BEJ \* BEJ XOC \* XOC KAZ \* KAZ RUW \* RUW GIY \* GIY POH \* POH WEF \* WEF QAM \* QAM ZIX \* ZIX NUK (Space 0, 2, or 8) \* NUK BEJ \* (etc.)

To break the chaining, the initial syllable in each of the 10 chained lists was replaced with a substitute syllable, i.e., LEB for ZIX in Chained 1; CIW for NUK in Chained 2; and TIV for BEJ, HUQ for XOC, VEP for KAZ, FOJ for RUW, LEQ for GIY, JAT for POH, YUS for

wef, and fup for qam, in the other 8 chained lists.

All syllables, including the substitute syllables, were from the Glaze 20% association value list.

*Procedure.*—The presentation rate was 1 sec. per syllable. Three spacing conditions were used: 0, 2, and 8 spaces, or seconds, between lists. The 10 lists, 3 spacing conditions, and 2 chaining conditions formed a 10 × 3 × 2 factorial design. Two Ss were assigned to each of the 60 experimental conditions.

The drum displayed an asterisk, followed by a stimulus syllable and then the response syllable. A stimulus syllable, therefore, appeared every 3 sec. After learning two 3-syllable practice lists in the same quasipaired-associates format to a criterion of one errorless trial, each S learned a 10-syllable list to a criterion of two successive errorless trials.

## Results

Spacing has the expected effect of decreasing total errors during learning. The effect is significant at the .01 level (see Table 3). The overall means of the converted errors for the zero-, two-, and eight-space conditions, in order, were as follows:



Fig. 3. Serial position curves with spacing varied: Exp. III.

13.6 $(SD = 2.1)$, 13.0 $(SD = 2.0)$, 12.0 $(SD = 1.7)$. The means of the number of trials to reach criterion were, in order, as follows: 44.7 $(SD = 18.2)$, 39.6 $(SD = 15.0)$, 31.9 $(SD = 11.1)$. Chaining does not have any significant overall effect. The serial position effect was significant $(P < .001)$, with the Greenhouse-Geisser lower bound test leaving the significance level unchanged.

The effect of spacing on the serial position curve is shown in Fig. 3. As spacing increases, the curve moves toward a markedly bowed shape of classic form, with the beginning lower than the end. The effect, as evaluated by the Position × Spacing interaction, is significant at the .001 level. Use of the Greenhouse-Geisser lower bound test with 2 and 60 df leaves the effect significant, at worst, at the .05 level. The effect of chaining on the serial position curve was not significant. The significant interaction of Lists and Serial Position $(P < .001)$ reflects differences in the difficulty of

### TABLE 3
ANALYSIS OF VARIANCE OF LOG ERROR SCORES: EXP. III

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | | | |
| Spacing (S) | 2 | 2.5434 | 6.09** |
| Chaining (C) | 1 | 0.0097 | |
| Lists (L) | 9 | 0.3966 | |
| S × C | 2 | 0.2680 | |
| S × L | 18 | 0.3114 | |
| C × L | 9 | 0.7316 | 1.75 |
| S × C × L | 18 | 0.2184 | |
| Error (between) | 60 | 0.4175 | |
| Within Ss | | | |
| Position (P) | 9 | 1.0312 | 56.35*** |
| P × S | 18 | 0.0609 | 3.33*** |
| P × C | 9 | 0.0250 | 1.37 |
| P × L | 81 | 0.0584 | 3.19*** |
| P × S × C | 18 | 0.0261 | 1.43 |
| P × S × L | 162 | 0.0172 | |
| P × C × L | 81 | 0.0178 | |
| P × S × C × L | 162 | 0.0184 | 1.01 |
| Error (within) | 540 | 0.0183 | |

** $P < .01$.
*** $P < .001$.

individual syllables, which were in different positions in the 10 lists.

The plots of the mean unconverted error scores display the same progression toward a more peaked serial position curve, with increased spacing. Plots of percentages of total errors, of course, give essentially the same picture, with respect to the shapes of the curves, as that obtained with mean log of errors. The data were also converted into ranks. Thus, if $S$ had the following raw error scores for the 10 positions: 6, 8, 15, 21, 18, 22, 19, 24, 13, 17, his scores were converted to the following ranks: 1.0, 2.0, 4.0, 8.0, 6.0, 9.0, 7.0, 10.0, 3.0, 5.0. Examination of the mean ranks for each position showed the same progression found with the mean log scores. Using the ranks, coefficients of concordance were computed for each group. For both the chained and unchained groups, the coefficients increased as spacing increased. In the chained group, they were .18, .30, .48; in the unchained group, they were .11, .42, .48. All the coefficients are significant ($P < .05$). The regular progression of the coefficients indicates that the increase in the bowing of the serial position curve as a function of spacing is independent of the particular conversion used.

Figure 3 shows a slight serial position effect in the zero-space condition. Separate analysis of the zero-space condition finds the effect significant ($F = 7.39, df = 1/60, P < .01$). The chained zero-space condition by itself also demonstrates a significant serial position effect ($F = 4.94, df = 1/60, P < .05$). The factor underlying this effect is the primacy-recency variable. Evidently, as in Mitchell's (1934) study, when spacing is absent, the primacy-recency factor affects the shape of the curve. The effect is, however, too slight to account for the usual serial position effect.

## DISCUSSION

The preceding experiments indicate that of the three factors listed initially as possible determinants of the serial position effect (chaining, spacing, and primacy-recency), the most important factor is spacing, the appearance of a gap in the list. Experiment I eliminated primacy-recency as a major factor in the usual serial learning situation, since a full, unaffected serial position effect was shown even when opposed by the primacy-recency factor. Experiment II gave some evidence that the serial position effect was a function of spacing. Experiment III completed the case in support of the effect of spacing and eliminated chaining as a major determinant of the serial position effect. Experiment III also indicated a slight effect for primacy-recency when it is not opposed by spacing.

McCrary and Hunter (1943), in their re-analysis of Hovland's (1938b, 1940) data, and Braun and Heymann (1958) display curves that remain unaffected by spacing, once the multiplicative effect of total number of errors is eliminated. Evidence has been presented here that spacing does have an effect. The apparent contradiction probably stems from differences in procedure. In the previous studies, the intertrial intervals were 6 and 126 sec.; in the present experiments, they ranged between 0 and 16 sec., with large changes appearing between the 0- and 2-sec. conditions. In the previous studies, an interpolated task was used, but only during the long interval. This could have counteracted the spacing effect by slowing the learning of the beginning and end syllables of the long interval lists.

Inhibition (Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940) and interference (Atkinson, 1957) constructs have been popular in the explanation of the serial position curve. From one point of view, the results above fit in with inhibition or interference explanations of the serial

position effect. Spacing could function as a barrier to protect the items near the beginning and the end of the list. With the demonstration of a systematic effect of spacing on the serial position curve, however, another type of explanation becomes possible: a facilitation explanation. With this type of explanation, interference or inhibitory effects between list items are considered homogeneous within the list. The serial position curve is viewed as a result of the facilitative effect of spacing on the learning of the first and last items of the list. These, in turn, facilitate the learning of their neighbors. The skewing of the curve can be explained by an effect demonstrated by Ribback and Underwood (1950). They showed that once an association is learned between a pair of syllables, it is easier to attach a third association to the second member of the pair than to the first. In other words, once the association X-Y is learned between syllables X and Y, it is easier to learn the triplet of syllables X-Y-Z than it is to learn A-X-Y. The Ribback-Underwood mechanism by itself is not sufficient to generate the serial position curve, since the mechanism does not determine the syllables from which it will spread. This determination could be made by spacing, which would facilitate the syllable pairs bordering the gap. These anchor syllables would in turn generate the serial position effect forward and backward from the gap.

## SUMMARY

Serial position effects are usually defined on the basis of the beginning of a list. In rote learning, the term "beginning" may be analyzed into three factors associated with a repetitively-appearing gap that separates the "end" from the "beginning" of the list: (a) Primacy-recency. The first item the S sees, and the last item he sees before the cycle repeats itself, appear on either side of the gap, because Es usually start the list from the gap. (b) Spacing. The appearance of the gap coincides with a period in which the S is not required to anticipate syllables. (c) Chaining versus association break. Every item in the series functions as both a stimulus and a response item (chaining), except the two on either side of the gap. One of these is solely a stimulus item; the other is solely a response item.

Three experiments were carried out to determine the role of these three factors in generating the serial position effect. In Exp. I, primacy-recency was opposed to both spacing and association break by varying S's starting position in conventional lists. No effect of the variation was found, indicating that primacy-recency was not a major factor. In Exp. II, spacing was varied, using intertrial intervals of 0, 2, 4, 8, and 16 sec. With the multiplicative effect of total number of errors held constant, some evidence was obtained indicating that the serial position curve became more peaked as spacing increased. In Exp. III, both spacing (intertrial intervals of 0, 2, and 8 sec.) and chaining or association break were varied. To vary chaining, the lists were given in quasipaired-associates form. Some lists required that an association be formed between the end and the beginning of the list; other lists did not. Experiment III showed clearly the effect of spacing in determining the serial position effect. There was no evidence for the effect of association break or chaining. Experiment III also indicated a slight effect for primacy-recency when it is not opposed by spacing. On the basis of these findings, it is concluded that the major factor determining the serial position effect is the amount of space between the end and the beginning of the list, with an increase in spacing producing a more marked serial position effect.

## REFERENCES

ATKINSON, R. C. A stochastic model for rote serial learning. *Psychometrika*, 1957, 22, 87–96.

BRAUN, H. W., & HEYMANN, S. P. Meaningfulness of material, distribution of practice, and serial-position curves. *J. exp. Psychol.*, 1958, 56, 146–150.

GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, 1959, 24, 95–112.

HOVLAND, C. I. Experimental studies in rote-learning theory: II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, 22, 338–353. (a)

HOVLAND, C. I. Experimental studies in rote-learning theory: III. Distribution of practice with varying speeds of syllable

presentation. *J. exp. Psychol.*, 1938, **23**, 172–190. (b)

HOVLAND, C. I. Experimental studies in rote-learning theory: VII. Distribution of practice with varying lengths of list. *J. exp. Psychol.*, 1940, **27**, 271–284.

HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.

McCRARY, J. W., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

MITCHELL, M. B. The effect of serial position in the continuous memorization of numbers. *Amer. J. Psychol.*, 1934, **46**, 493–494.

RIBBACK, A., & UNDERWOOD, B. J. An empirical explanation of the skewness of the bowed serial position curve. *J. exp. Psychol.*, 1950, **40**, 329–335.

ROBINSON, E. S., & BROWN, M. A. Effect of serial position upon memorization. *Amer. J. Psychol.*, 1926, **37**, 538–552.

# EFFECTS OF INSTRUCTIONS IN PROBABILITY LEARNING [1]

J. McCRACKEN, C. OSTERHOUT, AND JAMES F. VOSS

*College of Wooster*

The present paper[2] reports the results of two experiments designed to investigate probability learning as a function of instructions. The major instructional aspects studied were type of problem presented and information regarding sequential characteristics of the task. Studies of the role of instructions per se (Anderson & Grant, 1957) have not been extensive. However, instructions used in studies of two-choice, noncontingent probability learning that allowed for the possibility of $E_1$-$E_2$ patterns and permitted $S$ to respond on a trial-by-trial basis have yielded $A_1$ probabilities approximately equal to $E_1$ probabilities (e.g., Engler, 1958; Grant, Hake, & Hornseth, 1951; Neimark, 1956). On the other hand, studies which provided additional information regarding the nonpatterned $E_1$-$E_2$ trials yielded $A_1$ frequencies which exceeded $E_1$ probabilities (Morse & Runquist, 1960; Rubenstein, 1959). Experiment 1 was designed, therefore, to test the hypothesis that event prediction instructions and event prediction instructions with additional information regarding the sequential nature of the task yield $A_1$ probabilities equal to and in excess of the $E_1$ probability, respectively. In addition, Exp. 1 tested the hypothesis that

instructions which present no problem, but tell $S$ to engage in a conversation, yield nondifferential $E_1$-$E_2$ effects. A transfer condition was employed in order to determine the permanency of instructional effects.

Experiment 2 was designed to test the hypothesis that more specific instructions to avoid a trial-by-trial basis of responding and instructions to consider the task as essentially a discrimination problem yield $A_1$ maximization. Discrimination instructions were considered to be the extreme case of the lack of a trial-by-trial basis of responding since $S$ was told to consider the $E_1$-$E_2$ trials either in blocks of trials or as one two-event discrimination.

## EXPERIMENT 1

### Method

The $E_1$-$E_2$ probability of .70 −.30, randomized within 20-trial blocks, was used for both experiments. Each of two 33% association nonsense syllables (Hilgard, 1951), PIB and FAJ, occurred as the $E_1$ event for one-half of the randomly assigned $S$s. The $S$s were students at the College of Wooster.

*Original training (OT).*—Instructions for Group 1 essentially asked $S$ to engage in a conversation:

You have before you a list of two nonsense syllables. Using only the two words on this list, you and I are going to hold a conversation. To begin this conversation, you will say one of the words to me. I, in turn, will say one of the words to you. You will look over your list and say the word which you think is a reasonable statement at that time. We will continue in this manner until the end of the experiment. Remember, use only the two words before you for all of your contributions to the conversation.

[2] The notation used in this paper is as follows: $A_1$ and $A_2$ denote $S$'s predictions of events $E_1$ and $E_2$, respectively, where $E_1$ is the more frequent (.70) event and $E_2$ the less frequent (.30) event.

Instructions for Group 2 asked $S$ to predict a sequence of events:

> You have before you a card with two nonsense syllables listed on it. Using only these two words, you and I will hold a special type of conversation. You will begin by choosing one of these words and saying it aloud to me and I will then respond with one of the words. From this point on, you will try to predict the word I will say. If you are correct in your prediction, I will say that same word after you. We will continue in this manner (your prediction, my response) until the end of the experiment.
> Your task is merely to try very hard to predict correctly each separate word I shall say.
> Remember, use only the two words before you for all of your predictions and try hard to get each of your predictions right.

Instructions for Group 3 presented $S$ sequential information and asked $S$ to regard the situation as one problem. The first two paragraphs of the instructions for Group 2 were used, followed by:

> There is no pattern to discover. Your task is merely to get as many predictions correct as possible. Consider the entire experiment as one general problem and solve it so as to have the most correct you possibly can. It may be necessary to have some wrong predictions in order to best solve this general problem. Remember: treat the entire experiment as one general problem and attempt to solve it in order to get as many correct predictions as possible —even if you must have some wrong predictions to do so.
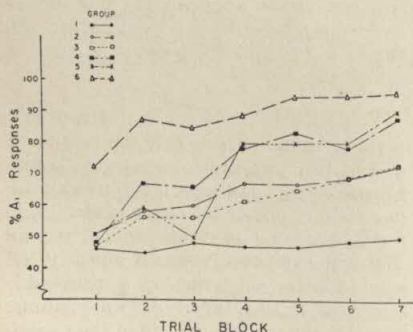


FIG. 1. Percentage of A₁ responses as a function of Trial Blocks 1–7 for Groups 1–6 of Exp. 1 and 2.

TABLE 1

PERCENTAGE OF $A_1$ RESPONSES FOR THE SEVEN 20-TRIAL BLOCKS OF THE NINE TRANSFER CONDITIONS

| Block Trial | Group | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 11 | 21 | 31 | 12 | 22 | 32 | 13 | 23 | 33 |
| 1 | 44 | 52 | 50 | 52 | 60 | 56 | 48 | 73 | 69 |
| 2 | 51 | 59 | 58 | 67 | 73 | 70 | 63 | 82 | 76 |
| 3 | 56 | 56 | 61 | 65 | 76 | 71 | 73 | 83 | 74 |
| 4 | 46 | 61 | 56 | 66 | 76 | 68 | 78 | 85 | 75 |
| 5 | 50 | 53 | 59 | 73 | 75 | 68 | 80 | 86 | 78 |
| 6 | 49 | 57 | 63 | 75 | 70 | 76 | 84 | 84 | 83 |
| 7 | 48 | 55 | 59 | 70 | 76 | 72 | 76 | 83 | 85 |

Note.—The first digit of the group number designates the OT condition, the second designates the T condition.

Thirty $Ss$ served in each group for 140 OT trials. The $E$ responded 2 sec. after $S$ gave a response.

*Transfer* $(T)$.—Each OT group was randomly divided into three equally sized subgroups. Each subgroup received one of the three sets of OT instructions after the 140 OT trials. The T (second instruction) groups were run for another 140 trials, with the same respective $E_1$-$E_2$ words, probability condition, and counterbalancing of words, but different 20-trial block randomized $E_1$-$E_2$ sequences.

## Results[3]

*Original training.*—Figure 1 presents the Groups 1, 2, and 3 $A_1$ percentages as a function of trial block for OT trials. Group 1 responded at approximately chance level, whereas Groups 2 and 3 approached matching. An analysis of variance performed on the $A_1$ frequency data revealed a significant instruction group effect ($F = 32.50$, $df = 2/84$). A Duncan range test indicated 1–32.[4] A Trial

[3] A preliminary experiment was performed involving Group 1 instructions and different $E_1$-$E_2$ sequences. The results were roughly equivalent to the present Group 1 results. The effect of the verbal counterbalancing and the interactions including word used are not significant for any analysis and are not further discussed.

[4] Significant differences are indicated by dashes between the group numbers, with an increase in $A_1$ frequency from left to right.

Block 7 analysis of variance revealed a significant instruction group source of variation ($F = 6.89$, $df = 2/84$) and the same range results.

*Transfer.*—Table 1 presents the percentage $A_1$ responses of the nine transfer groups for Trial Blocks 8–14. The data of Fig. 1 and Table 1 indicate a relatively rapid shift in performance level with the introduction of the T instructions. Analysis of variance performed on the $A_1$ frequency data of Trial Blocks 8–14 revealed a significant T instruction group source of variation ($F = 25.32$, $df = 2/72$). Subsequent range test results indicated 1-2-3. The OT instruction and T Instruction × OT Instruction sources of variation are not significant.

Analyses of variance were performed on the data of Trial Blocks 8 and 14. The OT instruction group source of variation in Trial Block 8 is significant ($F = 6.38$, $df = 3/72$), with subsequent range test results of 1-23. These results agreed with the results of Trial Block 7. The OT instruction group source of variation for Trial Block 14 is not significant. The Trial Block 8 and 14 findings likely reflect the reduced influence of OT instructions.

The T instruction group source of variation is significant in the Trial Block 8 ($F = 6.49$, $df = 2/72$) and Trial Block 14 ($F = 21.02$, $df = 2/72$) analyses. Subsequent range tests revealed: 12-3 and 1-23 on Trial Block 8 and 1-23 on Trial Block 14.

## EXPERIMENT 2

The lack of difference between Groups 2 and 3, except for the overall T analysis, implied that the sequential information and suggested approach presented to Group 3 did not differentiate the problem from the simple event prediction for Group 2. Experiment 2 was therefore designed to test the hypothesis that instructions to avoid a trial-by-trial basis of responding and instructions which force S to consider the task as essentially one two-choice discrimination problem yield maximization.

Group 4 was presented instructions to consider the task as one problem and to *avoid a trial-by-trial basis of responding*. Group 5 was instructed to predict one event when told to "Choose" (after 20 trials) and to maintain that response until again told to "Choose" (after 20 trials). Group 6 was instructed to consider the situation as a discrimination task and select the response more likely to occur once such a discrimination was made.

## Method

The first paragraph of instructions for Groups 4, 5, and 6 was the same as that of Groups 2 and 3. The Group 4 instructions continued:

Please note that you should ignore a trial-by-trial approach as much as possible. Neglect of this approach is helpful because there is no pattern or system to the sequence of words I shall say. Thus, you should approach this task as one problem and use any strategy you are able to devise in order to get as many words correct as possible. You will not be able to get all of the words correct, but you should try to get a maximum number of words correct. Remember, you are to approach the task as one problem and employ a strategy which will enable you to make a maximum number of correct predictions.

Group 5 instructions were:

Please note that you should ignore a trial-by-trial approach as much as possible. Neglect of this approach is helpful because there is no pattern or system to the words I shall say. In addition, you will not be able to get all of the words correct. However, in order to facilitate your getting a maximum number of correct predictions, the following procedure will be used. When we begin, you will try to predict the word I shall say. Then, at a given point in the sequence, I shall say "Choose."
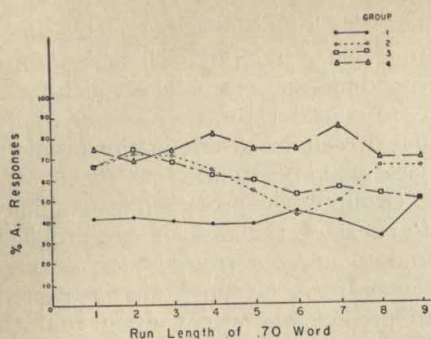
FIG. 2. Percentage of $A_1$ responses as a function of $E_1$ run length for Groups 1, 2, and 3 of Exp. 1 and Group 4 of Exp. 2.

After I say "Choose," you are to again predict, but you must predict with the same word until I again say "Choose." This procedure will continue until the end of the experiment. Remember, when I say "Choose," you must only predict with one of the words until I again say "Choose." Naturally, you should say the word that you think will occur more frequently until I again say "Choose" so that you will make a maximum number of correct predictions.

Group 6 instructions were:

Please note that you should ignore a trial-by-trial approach as much as possible. Neglect of this approach is helpful because there is no pattern or system to the words I shall say. In addition, I want to point out that you will not be able to get all of the words correct. However, in order to facilitate your getting a maximum number of correct predictions, you are to use the following procedure.

Consider the experiment as essentially one of discrimination. Therefore, if you are able to discriminate or tell which word occurs more frequently, you are to respond with that word on all of the trials. Since there is no pattern to the words, you will get as many words correct as possible by responding in this manner. Remember, as soon as you are able to, you should tell which word occurs more frequently and respond with that word *on each prediction*.

The $E_1$-$E_2$ sequences were those of Trials 1–140 of Exp. 1. The $N$ was 10 for each group. Other conditions were comparable to those of Exp. 1.

*Results*

The $A_1$ percentages for Groups 4, 5, and 6 during Trial Blocks 1–7 are presented in Fig. 1. A test for significance between Groups 4, 5, and 6 was not performed because of the limitations placed upon the $A_1$ responses by the instructions.

The results of Group 4 indicate that instructions to respond to the situation as one problem and to avoid a trial-by-trial approach yielded an $A_1$ percentage approaching 90%. The Group 5 $A_1$ percentage approached approximately the same level. Group 6 tended to choose the more frequent word during the first trial block, although after an $E_2$ run, an $A_2$ response occasionally occurred.

The differences among Groups 1, 2, 3, and 4 were clarified by an analysis of negative recency effects. Figure 2 presents the percentage of $A_1$ responses as a function of $E_1$ repetitions. Groups 1 and 4 did not show negative recency effects, whereas Groups 2 and 3 yielded such tendencies. These results imply that Groups 2 and 3 responded more to sequential characteristics of the stimuli than Groups 1 and 4. The frequency distributions of $A_1$ responses of the Trial Block 7 data of Groups 1, 2, 3, and 4 were essentially unimodal.

### DISCUSSION

In general, the findings indicated that conversational instructions yielded chance results; event prediction instructions, with or without additional sequential information and problem orientation, yielded approximate matching; event prediction instructions which included the avoidance of a trial-by-trial orientation yielded an $A_1$ frequency greater than the $E_1$ frequency; and instructions which forced $S$ to choose one alternative for blocks of 20 trials or to discriminate which event occurred more frequently yielded asymptotic $A_1$ levels of approximately 90% to 95%. The results of Group 5 are in agreement with the results of Galanter and Smith (1958).

The results are relevant to the area o

decision making in a probabilistic situation (Edwards, 1961) in that the $A_1$ response level is a function of the problem presented. Instructions which include the possibility of $E_1$-$E_2$ patterns and do not specifically state that a trial-by-trial approach should be avoided apparently yield results based upon sequential $E_1$-$E_2$ characteristics. On the other hand, the results of Groups 5 and 6 suggest that maximimizing occurs when the task is presented as one discrimination-type problem and the responses indicated are $A_1$ and $A_2$ for large blocks of trials rather than trial-by-trial sequences of $A_1$ and $A_2$ responses.

The conversation instruction results suggest that in a probability situation, differential $E_1$-$E_2$ frequency is not a sufficient condition to yield $A_1$-$A_2$ differences.

## SUMMARY

Instructions in two .70 − .30 noncontingent probability experiments were varied for six groups. Groups 1 to 6 were told to consider the task as (a) conversation, (b) event prediction, (c) event prediction—with additional task information, (d) event prediction— avoiding a trial-by-trial basis of responding, (e) a problem involving responding for blocks of 20 trials, and (f) a problem involving the discrimination of the two events.

Group 1 responded at an approximate 50% $A_1$ level, Groups 2 and 3 approximately matched, and Groups 4, 5, and 6 exceeded 85% $A_1$ responding. Analysis of negative

recency effects for Groups 1–4 suggested that Groups 1 and 4 were not responding to the sequential nature of the task, whereas Groups 2 and 3 were.

## REFERENCES

ANDERSON, N. H., & GRANT, D. A. A test of statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, 54, 305–317.

EDWARDS, W. Behavior decision theory. *Annu. Rev. Psychol.*, 1961, 12, 473–498.

ENGLER, J. Marginal and conditional stimulus and response probabilities. *J. exp. Psychol.*, 1958, 55, 303–317.

GALANTER, E. H., & SMITH, W. A. S. Some experiments on a simple thought problem. *Amer. J. Psychol.*, 1958, 71, 359–366.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, 42, 1–5.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 517–567.

MORSE, E. B., & RUNQUIST, W. N. Probability-matching with an unscheduled random sequence. *Amer. J. Psychol.*, 1960, 73, 603–607

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, 52, 209–220.

RUBENSTEIN, I. Some factors in probability matching. *J. exp. Psychol.*, 1959, 57, 413–416.

# COGNITIVE FACTORS IN HEART RATE CONDITIONING [1]

## BISHWA B. CHATTERJEE [2] AND CHARLES W. ERIKSEN

### University of Illinois

Studies on autonomic conditioning in human $S$s typically have given little attention to the role of $S$'s cognitive or verbalizable expectancies upon the development of the conditioned response. Where attention has been directed to these variables the emphasis has been on demonstrating that humans can develop conditioned autonomic responses without verbalizable awareness. Experiments by Diven (1937) and Haggard (1943) have been interpreted as showing that human $S$s will condition GSRs without verbalizable awareness of the contingent relationship between the CS and the UCS and Razran (1946) has interpreted his data on salivary conditioning along the same lines.

In these studies, however, only a crude assessment was made of $S$'s verbalizable awareness. There was no attempt to make a fine grained analysis of specific expectancies and their relation to autonomic response. Even the conclusion of conditioning without awareness is extremely equivocal. Lacey and Smith (1954) have shown that neither the Diven nor Haggard studies contain the necessary logical or statistical comparisons of data to permit a conclusion of unconscious conditioning. Further, Chatterjee and Eriksen (1960) have shown that Lacey and Smith's own data on heart rate conditioning without awareness may well have been due to an artifact arising from the method of computing conditioning scores.

There are but few studies that have

attempted to manipulate experimentally $S$'s cognitive expectancies in autonomic conditioning or to correlate closely his verbalizable expectancies with autonomic responses. Notterman, Schoenfeld, and Bersch (1952) informed one group that there would be no further shocks and found extinction of conditioned heart rate was much quicker than in an uninformed group. Branca (1957) undertook intensive questioning of his $S$s and related their expectations of shock as well as their verbalizations as to whether the shock was painful or not to the frequency and occurrence of conditioned GSRs. He concluded (Branca, 1957) "expectation of shock as a painful or fearful experience was necessary and sufficient to produce responses to the experimental and generalization stimuli in this experiment and such expectancy was the result of awareness of the existing relationships between the experimental stimuli and experience with the unconditioned stimulus" (p. 549).

In the present experiment we studied the acquisition and extinction of a conditioned heart rate response under different conditions of experimentally manipulated cognitive expectancies. Since previous methods of determining autonomic conditioning have been shown to result in spurious evidence of precise conditioning (Chatterjee & Eriksen, 1960; Eriksen, 1958) an improved and more rigorous criterion of conditioning was employed. In addition the generalization of conditioned heart rate responses along a semantic and a color dimension was determined.

## METHOD

*Design.*—Stimuli were common words presented to *S* by means of an optical projector upon a ground glass screen at exposure duration of .5 sec. The *S* began chain associating to the word upon exposure. After 8.5 sec. a small light located above the screen flashed for 1 sec. This was the signal for *S* to stop associating. Following a 15-sec. rest interval, the next stimulus word was presented. The stimulus list consisted of 12 nouns and adjectives all with a high frequency count in the Thorndike-Lorge (1944) tables.

In the conditioning session the stimuli were arranged in seven blocks of trials, each block containing all 12 words randomized within the block. In the extinction phase, which followed the conditioning phase without interruption, a second list of 12 words was used. Only one word, BOAT, the CS, was common in both lists. The extinction session consisted of three blocks of trials with each block again containing randomization of the 12 words. Each word was exposed with a deep colored background of either green, yellow, or red. There were four words in each color in both stimulus lists. The extinction list contained the word BOAT, five words semantically related to BOAT, and six words semantically dissimilar. The stimulus words with the associated colors for the conditioning and extinction phases are shown in Table 1.

To establish cardiac conditioning a painful shock was always presented at the termination of chain association to the word BOAT. The shock, lasting for 1 sec., coincided in time and duration with the light signal for stopping associations. There were no shocks in the extinction phase.

The electrotachographic response of *S* was continuously recorded during the experimental period. Upon completion of the extinction trials *S* was interrogated with a series of questions to estimate his awareness of shock frequency following each of the stimulus words.

*Procedure.*—The *S* was seated in a comfortable reclining chair facing a large wood screen that contained the ground glass projection screen. All of the apparatus was concealed from the sight of *S*. The *S* was instructed as follows:

We're going to do an experiment on chained association. A common word will be flashed for one-half second on this screen and you are to read that word and then go on speaking the words that come to your

## TABLE 1

COLOR SCHEME OF THE STIMULUS WORDS

| Conditioning Stimulus List | | Extinction Stimulus List | |
|---|---|---|---|
| Word | Color | Word | Color |
| BOAT | green | BOAT | green |
| CHILD | green | SEA | green |
| RICH | green | FACE | green |
| PAPER | green | INDUSTRY | green |
| BIRD | yellow | SHIP | yellow |
| HOUSE | yellow | SAIL | yellow |
| ALONE | yellow | DRESS | yellow |
| BLUE | yellow | PLOT | yellow |
| SOFT | red | VESSEL | red |
| TREE | red | WAVE | red |
| SKY | red | NATURE | red |
| PEACE | red | MONEY | red |

mind. Do not try to make any meaningful sentences or any sense out of your free association. Rather, give your mind a free rein. It is best to be relaxed and not hold back anything. Continue speaking one word after another until you notice this little light bulb flash for one second. This is the signal for you to stop. Then wait until the next word is flashed on the screen when you again begin your free associations as before. We will use only a limited number of words over and over again.

The *S* was then given training with a few practice words and any questions he had were answered. Heart rate was recorded through Standard Lead 1 and shock was administered through electrodes placed on the right calf with 1½-in. interelectrode distance. The *S* was told that electric shock was an integral part of the experiment and that we would begin by determining his threshold for pain. A number of shocks were then administered at increasing voltage levels until *S* reported the shock as quite painful. This was the level used for that *S*.

Three basic groups of *S*s were formed by adding to the above instructions in the following ways. Group I *S*s were told that a shock would follow one particular word but that no other words would be followed by shock. They were further told when the extinction trials began by informing them there would be no further shocks. Group II *S*s were told that following the presentation of a particular word in the list there would always be a shock and each of the remaining words in the list

would be followed by one shock sometime during the presentation. They were further told that during the latter part of the experiment all the shocks would cease. Group III was told that a certain number of shocks would be administered at certain points of time during the experiment but $E$ could not tell them beforehand when the shocks would come. These differences in instructions for the three groups were designed to lead to different cognitive expectancies concerning conditioning arrangements. In keeping with these differences in instructions Group I $Ss$ were given only 7 shocks while $Ss$ in Groups II and III received 18 (7 to the word BOAT and 1 each to the remaining 11 words present during the conditioning trials).

Before beginning the actual conditioning phase an adaptation period was employed in which 12 training words were used.

Upon termination of the extinction phase the electrodes were removed and all $Ss$ were put through an interrogation as to their expectancies and beliefs concerning the nature of the experiment. In addition to questions designed to elicit $S$'s hypotheses as to the purpose of the experiment and the basis on which shock was administered, the list of stimulus words were administered to $S$ one by one and he was asked to indicate the number of times he thought shock had followed each. This provided an estimate of $S$'s expectancy of shock for the different words.

*Apparatus.*—The stimulus words were typed in small letters and mounted in 35-mm. slides. Colored cellophane sheets were wrapped around the slides to provide variation in color. A Bell and Howell Robomatic slide projector was used in projecting the stimuli on the ground glass screen. Stimulus presentation, shock administration, and duration were automatically programed and timed.

A Grass Model 5P4 polygraph was used for continuous recording of $Ss$' electrotachographic responses. In addition to the pen recording the tachogram, three event marking pens were also triggered from appropriate leads from the timing equipment. They were used to indicate the onset and termination of the CS, of the stop signal, and the UCS.

Shock was faradic stimulation from a Harvard inductorium supplied with a 3-v. dc source in the primary.

*Subjects.*—Seventy-six undergraduate students (19 females and 57 males) in introductory psychology and education courses were used. Assignment to groups was alternated on the basis of the order of their appearance for the experimental session

except that the number of females in each group was made roughly proportionate.[3]

## RESULTS

The response measure used in assessing conditioning, extinction, and generalization was a cardiac response difference score. The latencies between the four heat beats immediately prior to the presentation of the stimulus were determined for each of the 120 stimulus presentations during the conditioning and extinction trials for each $S$. Also, the latencies between the four beats immediately following presentation of the CS were determined. The cardiac response difference was obtained by subtracting the shortest post-CS latency from the shortest pre-CS latency on each presentation.[4]

The 9 $Ss$ in Group I were all able to verbalize during the inquiry period that shock had followed only the critical word BOAT. The 11 $Ss$ constituting Group II had received a total of 18 shocks: 7 to the word BOAT and once each to the remaining 11 words during the conditioning phase. Again all these $Ss$ reported during the inquiry that they had discriminated the word BOAT from the remaining words by the third or fourth presentation of BOAT. On the other hand, none of the other 56 $Ss$ comprising Group III were able to verbalize clearly and unequivocally that shock had followed BOAT nearly all the time during the conditioning trials and occurred only

[3] After 11 $Ss$ had been run in each group, the data were examined. Due to recording failures 2 $Ss$ had to be discarded from Group I. The existence of conditioning in Groups I and II was at this point quite evident, but the negative results in Group III led us to add an additional 45 $Ss$ to this group in order to insure a more sensitive test of this condition.

[4] The cardiac UCR in the present experiment was almost without exception a decrease in latency between beats.

once to each of the remaining words. However, when Ss were asked to state their expectancies for shock to each of the stimulus words, it was possible to subdivide Group III into two groups: Group III-A, the partially aware group consisted of 29 Ss who met the criterion of reporting that they had received at least three or more shocks to the word BOAT and not more than two shocks to any of the remaining words. The remaining 27 Ss formed Group III-U. These Ss reported frequency of shock to the word BOAT less than or equal to two times or reported more than two shocks to BOAT but at the same time reported an equal or greater number of shocks to one or more of the remaining words. This division of Group III was employed to permit a more refined analysis of the relationship between cognitive expectancy and conditioning and generalization.

*Conditioning and extinction.*—Since conditioning is a monotonic increasing function of the number of reinforced trials, a rigorous test of conditioning is to plot the number or percent of Ss in each experimental group who gave the maximum positive cardiac response difference to BOAT in each block of trials (chance level is 8.3%).

The results of this measure of conditioning and extinction are plotted in Fig. 1. It is apparent that the conditioning varies from group to group. It is greatest in Group I, somewhat less so in Group II, and in Groups III-A and III-U the curves indicate that no conditioning occurred.

The binomial test was used to determine whether the obtained proportions of Ss in each group giving the maximum response to BOAT on the last acquisition trial block differed significantly from the chance level. For Group I the probability is .003;

for Group II, .06; for Group III-A, .11; and .40 for Group III-U.

The above method for assessing conditioning is quite rigorous and is designed to reveal the specificity of the stimulus-response relation. It is informative, however, to examine conditioning and extinction in the different groups in terms of an alternative criterion. First, the sizes of the response differences in each block for each word were ranked separately from small to large for each S and then the rank value for BOAT was averaged through Ss by group and trial block. Curves obtained in this way are shown in Fig. 2.

While this measure is more comparable to the averaging techniques that have been used in previous autonomic conditioning studies it nonetheless reveals essentially the same conclusions as the previous analysis. Again Groups I and II show evidence of conditioning whereas Groups III-A and III-U do not.

An analysis of variance following
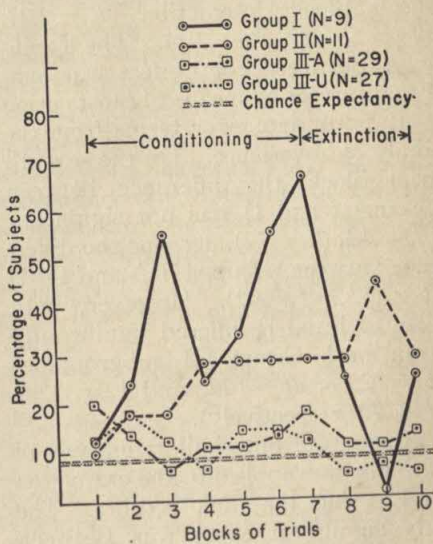


FIG. 1. Percentage of Ss whose maximum cardiac response difference follows the stimulus BOAT.
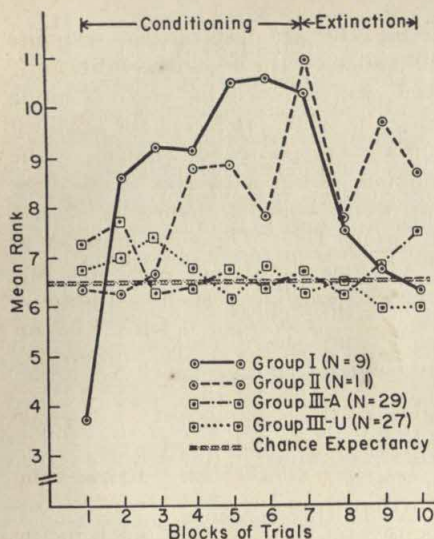
FIG. 2. Mean rank of cardiac response difference which follows the stimulus BOAT.

Lindquist's (1953) Plan I for the conditioning trials only was used to test the significance of the effects in Fig. 2. There was a significant Between-Group effect ($F = 6.35$, $df = 3/72$, $P < .01$) and a significant Group × Trial Blocks interaction ($F = 2.27$, $df = 18/432$, $P < .01$). The significant Between-Group effect was analyzed further by means of $t$ tests using the appropriate error terms from the analysis of variance. On the seventh trial block, the difference between Groups I and II was not significant ($t = .14$, $df = 72$), nor was the difference between Groups III-A and III-U ($t = .55$, $df = 72$). However, both Groups I and II differed significantly from both Group III subgroupings ($t = 2.88$, $df = 72$, and $t = 2.95$, $df = 72$, respectively).

A separate but similar analysis of variance was applied to the extinction trials using the rank scoring. The only significant effect was Between Groups ($F = 3.32$, $df = 3/72$, $P < .05$). Further analysis by the $t$ test revealed the significant Between-Group effect

to be due to the differences between Groups II and III-U.

*Cardiac responsiveness.*—While the above results are quite clear in indicating the role of cognitive expectancies on the acquisition of conditioned heart rate responses, it is also informative to investigate the effects of these cognitive expectancies upon other characteristics of heart rate behavior, specifically heart rate responses to nonconditioned words. This was done by determining the cardiac response difference to all words except BOAT for each trial block for each $S$. In Fig. 3 the average response difference for these 11 words has been plotted as a function of trial blocks during conditioning and extinction for each of the four groups.

There is a tendency for the cardiac response difference to words other than the CS to decrease throughout the conditioning and extinction trials. Also there is a marked difference be-
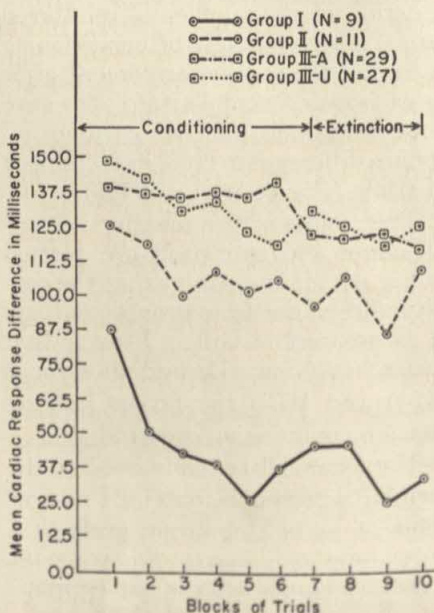


FIG. 3. Mean cardiac response difference following all stimulus words excluding BOAT.

tween groups. Group I shows less reactivity to nonconditioned words throughout the conditioning and extinction sessions with Group II falling between Group I and Groups III-A and III-U. The latter two groups are indistinguishable in performance.

The significance of the effects shown in Fig. 3 was evaluated by an analysis of variance using Lindquist's Plan I. In this analysis conditioning and extinction trials were included in the same analyses. The only significant effects were for Between-Trial Blocks ($F = 4.13$, $df = 9/648$) and Between Groups ($F = 10.03$, $df = 3/72$).

Since the biggest group difference in Fig. 3 is due to Group I, it was desired to determine whether the different cognitive expectancies between Group II and Groups III-A and III-U had an effect upon the cardiac response difference to the nonconditioned stimuli. To determine this, the above analysis was repeated using only the latter three groups. Again there was a significant Trial Block effect ($F = 3.90$, $df = 9/676$), but the Between-Group effect ($F = 1.41$, $df = 2/64$) was not significant.

The above analyses have suggested that heart rate conditioning and heart rate behavior are in part a function of $Ss$' verbalized expectancies concerning shock. We can ask the reverse question as to whether differences in heart rate behavior will predict $Ss$' verbalizations. For this question the data of Group III are available. The $Ss$ in this group were asked to report the number of shocks they thought they had received to each of the stimulus words. Our previous analysis would suggest that words having a large cardiac response difference should have a higher number of reported shocks than words with a low cardiac response difference. To test this possibility the word

giving the greatest cardiac response difference on the seventh conditioning trial was selected for each $S$ along with the word producing the smallest cardiac response difference. The number of reported shocks to these two words was determined for each $S$ in Group III. The mean number of shocks reported to the word with the greatest cardiac response difference was 1.37 as compared with 1.03 for the word with the smallest difference. A $t$ test for correlated scores gave a value of 1.71 significant at the .05 level for a one-tailed test.

*Tests of generalization.*—In assessing generalization only the cardiac responses on the first trial block in the extinction series were examined. As will be recalled, the stimulus words for extinction contained not only the CS but words semantically similar and dissimilar and colors that were the same and different from the CS.

Detailed analyses of the data from Group III failed to yield any significant or suggestive evidence of either semantic or color generalization. Similarly, separate analyses of the data from Groups I and II who had shown conditioning, gave no evidence of generalization. Since the previous analyses had suggested that cognitive expectancy was an essential correlate of heart rate responses, and these $Ss$ had not verbalized semantic or color relations, the lack of generalization might have been anticipated.

## DISCUSSION

Conditioning was evident in Groups I and II but here all $Ss$ had clearly verbalizable expectancies concerning the relationship between stimulus and shock. The heart rate response of $Ss$ in these two groups shows a further correspondence with their cognitive expectancies in the extinction behavior. Group I $Ss$ were told following the seventh conditioning

trial that there would be no further shocks. The data are clear in showing almost complete extinction on the first extinction trial. Group II Ss on the other hand were not informed of the termination of the conditioning trials and there is little or no evidence of extinction of their heart rate response during the three extinction trial blocks.

Group III Ss who had received a minimum amount of information concerning the relationship between the stimuli and shocks prior to the conditioning session show no evidence of having conditioned heart rate responses. Even when this group of Ss is subdivided into subgroups based upon their verbal expectancies of shocks obtained postexperimentally there is no evidence that the subgroup showing some verbal discrimination between the shock and the nonshock stimuli has conditioned. It may be that it was only in the last of the conditioning trials that these Ss began to form cognitive expectancies concerning the relationship between specific stimuli and shock. Thus there was insufficient time for conditioned heart rates to occur.

But even among the Group III Ss there is some evidence of a relation between cognitive expectation and heart rate behavior. When the stimulus word with the greatest heart rate response is compared with the smallest heart rate response on the last conditioning trial, it is found that these words are discriminated in the terms of S's verbalized shock expectancies.

The effects of cognitive expectancies is also apparent on other aspects of heart rate behavior in the experimental situation. Group I Ss who knew that only one word would be shocked during the experimental session showed an appreciably smaller heart rate response to the nonconditioned stimulus words throughout the conditioning and extinction session. Similarly Group II Ss who knew that words other than the CS would be shocked only once during the session showed less heart rate response to these nonconditioned stimuli than did the Group III Ss who had no definite expec-

tation of relationships between stimuli and shock.

The absence of any stimulus generalization along either the semantic or the color dimension is also consistent with the findings concerning the importance of cognitive expectancies on the occurrence of heart rate response changes to the stimuli. Since conditioning occurred only in the Group I and II Ss generalization can only be expected to occur in these groups. However, Ss in these groups had clearly verbalizable expectancies concerning the relationship between the conditioned stimulus and shock. Since their verbalizations during the inquiry period did not express relationships between colors or semantic class and the occurrence of shock, the hypothesis that cognitive expectancy determines heart rate behavior would predict a failure of generalization to occur among these Ss.

While the data considered so far are quite unequivocal in demonstrating a relationship between heart rate behavior and cognitive expectancy the question can be raised as to whether the cognitive expectancy is necessarily prior to or a determiner of the heart rate phenomena. There are several factors in the present experiment that would indicate that cognitive expectancy is a determiner of the heart rate response rather than something that develops concurrently with the conditioned heart rate. In previous studies the typical procedure has been to allow awareness or expectancies to develop along with the CR. In this type of design it is impossible to determine causal relationships. However, by manipulating the expectancies of our Ss prior to the beginning of the conditioning sessions we have largely controlled the expectancies that existed in our Ss before conditioning of the heart rate occurred.

There is further evidence of the primacy of cognitive expectancies in the extinction behavior of the Group I Ss. The knowledge that no further shocks would occur was sufficient to produce almost complete extinction without experiencing the CS in the absence of the

UCS. This result is consistent with that of Notterman et al. (1952).

The results we have obtained are also quite consistent with those of Branca (1957), who reports marked correspondence between conditioned autonomic behavior and Ss' verbalizable expectancies in the conditioning situation.

One further comment is in order concerning the low relationship between cognitive expectancy and heart rate behavior. While the above evidence shows a definite relationship between cognitive expectancies and heart rate behavior the relationship is certainly not a very high one. Eriksen (1958) has pointed out elsewhere the existence of large noncorrelated errors between different response systems such as the verbal and the autonomic and the present low relationships are probably a reflection of the amount of error in cognitive expectancies as well as heart rate behavior. In view of other evidence it is most likely that the largest source of error is in the heart rate response rather than in Ss' verbalization of his expectancies.

## SUMMARY

The conditioning and semantic and color generalization of the heart rate was studied as a function of different cognitive expectancies of the Ss. Cognitive expectancies were manipulated prior to conditioning by means of different instructions to the three experimental groups.

A high correspondence was found between heart rate and verbalizable expectancies. Clear evidence of heart rate conditioning was obtained only in those cases where S could verbalize the relationship between the CS and UCS. Those Ss who were informed that there would be no more shocks at the beginning of the extinction trials showed almost a complete loss of the CR without experiencing nonreinforced presentations of the CS. Heart rate responses to nonconditioned stimuli were also found to vary as a function of cognitive expectancies. There was some indication that observed differences in heart rate could be used to predict differences in verbalized expectancies.

There was no evidence of either semantic or color generalization of the conditioned heart rate. This finding was considered consistent with the above since Ss did not include such generalized expectancies in their verbalizations.

## REFERENCES

BRANCA, A. A. Semantic generalization at the level of the conditioning experiment. *Amer. J. Psychol.*, 1957, **70**, 541–549.

CHATTERJEE, B. B., & ERIKSEN, C. W. Conditioning and generalization as a function of awareness. *J. abnorm. soc. Psychol.*, 1960, **60**, 396–403.

DIVEN, K. Certain determinants of conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 169–227.

HAGGARD, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomous reactions during and following stress. *J. exp. Psychol.*, 1943, **33**, 257–284.

LACEY, J. L., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

NOTTERMAN, J. M., SCHOENFELD, W. N., & BERSCH, P. J. Partial reinforcement and conditioned heart rate response in human subjects. *Science*, 1952, **115**, 77–79.

RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1946, **46**, 337–365.

THORNDIKE, E. L., & LORGE, I. *Teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

# TEST OF THE HYPOTHESIS OF PSYCHOLOGICAL REFRACTORY PERIOD [1]

## JACK A. ADAMS

### University of Illinois

Telford (1931) found that the second of two reaction time measures was lengthened when the spacing of two successive stimuli was reduced to 500 msec., and his hypothesis of psychological refractory period has motivated a number of confirming studies and analyses (Adams, 1961; Craik, 1948; Davis, 1956, 1957, 1959; Marill, 1957; Vince, 1948; Welford, 1952, 1959). A prominent interpretation of these findings is that an incoming stimulus is subjected to a central decision process before discharge occurs down the motor nerves, and a second stimulus impinging during this decision time is either disregarded, degraded, or delayed in immediate memory until the decision mechanism has been cleared. A general conclusion is that $S$ is a one-channel data processing system.

Hick (1948), Poulton (1950), and Elithorn and Lawrence (1955) have suggested a counterexplanation in terms of expectancy. In this usage, "expectancy" refers to $S$ having learned certain properties of the statistically defined time relationships between the first and the second stimulus presented over a relatively long series, and $S$ is thought to be most alert for responding when the interstimulus interval is somewhere in the vicinity of the mean delay. When a very short interval occurs, $S$ is not

ready, and his response to the second stimulus is lengthened. An implication of the expectancy hypothesis is that $S$ can be a multichannel system when conditions allow the acquisition of appropriate expectancies.

All findings can be explained about equally well by both hypotheses, but a discriminating test seems possible by manipulating the statistical structure of interstimulus time intervals. So far, studies of this topic have used only highly uncertain, random interval distributions (low redundancy). The expectancy hypothesis would predict that refractoriness is a function of the statistical structure of time intervals by making $S$ more expectant for certain classes of intervals, while the one-channel hypothesis would regard refractoriness a function of intervals but not statistical structure. The investigation reported here manipulated the redundancy of interstimulus intervals in a two-dimensional, bisensory discrete tracking task where an audio signal occurred with or lagged a visual signal by a defined time interval.

## METHOD

*Apparatus.*—The discrete tracking apparatus was used (Adams & Chambers, 1962). This device can be used as a two-dimensional bisensory discrete tracking task where the visual and the audio inputs each have three discrete states, or as a one-dimensional tracking task with either audio or visual stimuli. In the visual dimension, $S$ had three horizontally arranged jeweled stimulus lights (red, white, and green) in front of him at eye level, and these lights came on in a repetitive sequence at defined time intervals (see below). Beneath each stimulus light was a small neon

response feedback light that informed $S$ of the position of his control, and this direct display of response feedback cues, as well as stimulus lights, made the task one of pursuit tracking. The $S$ had to keep the feedback light aligned with the frequently changing stimulus light as much as possible. The audio dimension had a 600-, 800-, and 1000-cycle pure tone as stimuli, and $S$ heard them over a headset and responded with the same type of control as used for the visual dimension. The auditory error coding was in terms of a complex tone. When $S$ had the control in the correct position, he heard a pure tone but, when he was wrong, he heard a complex tone made up of two of the three fundamental frequencies. The correct stimulus was presented as a pure tone in the complex, and, superimposed on it was a second tone, rapidly interrupted, whose frequency was determined by the position of the control. Thus, when $S$ was in error he had feedback on the correct stimulus and the wrong control position, and this was formally equivalent to the pursuit tracking format of the visual task. Moreover, being pursuit tracking, $S$ had direct feedback of his response time to stimulus change in each dimension. He could see how long it took for the visual feedback light to become aligned in the correct position, and similarly he could hear the duration of the interrupted tone. The instructions were to always respond and nullify error as quickly as possible and, in the bisensory task, to give equal attention to both dimensions.

Response was with a 2-in. control stick mounted on the wide armrest of $S$'s chair, and he could have a control for one or both hands, depending on whether tracking was unisensory or bisensory. The control moved freely through its arc, although electrically it could only assume three states. Audio or visual stimuli could be switched to either control. The direction of control movement was always horizontal. An Esterline-Angus operations recorder allowed $E$ to completely record all stimulus and response events and their time relationships on a trial.

*Procedure.*—There were three groups of $S$s, distinguished by the statistical distribution of time delay intervals governing the amount that an audio stimulus followed a visual one in two-dimensional bisensory tracking. Each group was given three practice sessions on different days, usually within the same week. A session was 12 2-min. trials, with 5 min. rest between Trials 4 and 5, and 8 and 9, and 50 sec. intertrial rest for the remainder of the trials. The basic phenomenon of psychological refractory period should be manifest

as delay in response to the audio stimulus when it follows the visual stimulus too closely, and it is necessary to have a unisensory audio control condition to demonstrate this delay. In addition, it was considered advisable to have a unisensory visual control condition to see if time uncertainty influenced response to the visual stimulus as well as audio. Each $S$ provided his own unisensory control measures. For the 12 trials of a session, 4 were unisensory visual, 4 unisensory audio, and 4 bisensory. The order of these three task conditions was counterbalanced among $S$s of each group, and the particular order assigned to an $S$ was the same on each session. Also, within each group, the assignment of visual and audio signals to left and right hands was counterbalanced. The operations recorder was used on Trials 2 and 3 of each block of 4 trials in Session 3 to provide a detailed analysis of individual responses.

*Stimulus series.*—The type, duration, and interstimulus intervals of audio and visual events were programed on punched tape and automatically read by a motorized tape reader. A single stimulus input tape was constructed for each group, and it had 60 audio-visual bisensory stimulus pairs on each trial. A given group used the same tape on each session. Whenever a unisensory trial was required, $E$ disconnected the unwanted stimulus series in the other sensory dimension from the presentation. The 60 visual events on a trial had durations of 1.5, 2.0, and 2.5 sec., and there were 20 of each duration. The order of the 60 visual events was separately randomized for each trial. All groups had the same order of visual events on a given trial.

The audio delay intervals were in the range where the phenomenon of psychological refractory period was expected to be maximal, as well as somewhat beyond this range, and were 0, 100, 200, 400, and 800 msec. The approach was mainly to manipulate the frequency of audio delay intervals of 100 msec., where the phenomenon of psychological refractory period is known to be high. If expectancy is a significant explanatory mechanism for behavior, the amount of degradation in audio RT should be influenced by the frequency of events at the very small delay intervals. The $S$ should be more expectant with a greater frequency of small intervals and consequently should have less decrement in audio RT. On the other hand, if expectancy is not a relevant explanatory framework, no difference should be expected. The statistical distribution of audio time delay intervals on a trial for each group, mean delay, and percent-

TABLE 1

NUMBER AND PROBABILITY OF DELAY
INTERVALS BETWEEN VISUAL AND
AUDIO STIMULUS ON A BISENSORY
TRIAL FOR EACH GROUP

| Audio Delay Interval (in Msec.) | Group | | | | | |
|---|---|---|---|---|---|---|
| | LU | | MU | | HU | |
| | $p$ | N | $p$ | N | $p$ | N |
| 0 | .05 | 3 | .10 | 6 | .20 | 12 |
| 100 | .80 | 48 | .60 | 36 | .20 | 12 |
| 200 | .05 | 3 | .10 | 6 | .20 | 12 |
| 400 | .05 | 3 | .10 | 6 | .20 | 12 |
| 800 | .05 | 3 | .10 | 6 | .20 | 12 |
| Mean (msec.) | 150 | | 200 | | 300 | |
| % Redundancy | 52 | | 24 | | 0 | |

age redundancy (Attneave, 1959) are shown
in Table 1. The groups were designated low
uncertainty (LU), medium uncertainty (MU),
and high uncertainty (HU). The order of
the delay intervals was separately randomized
for each trial.

It should be emphasized that the experi-
ment deals only with time uncertainty in
tracking, not event uncertainty as investi-
gated by Adams and Chambers (1962) with
this task. Each sensory dimension was a
repetitive series of the three stimulus events
which simply required S to move the control
back and forth. And, on the bisensory trials,
the same audio and visual events were always
paired together. Thus, the stimulus series
always had *event certainty*. However, because
the time patterning of events was statistically
determined, the stimuli always had *time un-
certainty*, which was the basic experimental
variable.

*Subjects.*—There were 18 Ss in a group.
The 54 Ss were university male under-
graduates who were paid for their participa-
tion. They were randomly assigned to
groups.

*Performance measurement.*—Overall pro-
ficiency in discrete tracking, such as mea-
sured by time on target, is a function of (a)
off-target time between the onset of the
stimulus and the onset of the response,
whether the response is correct or not, (b)
number of errors, i.e., movements of the
control to the wrong position, and (c) duration
of each error before it is corrected. Because
the motor movements required were simple

and repetitive, errors were negligible and so
proficiency primarily was determined by the
off-target time of a correct response to change
of a stimulus light or tone. These time values
for individual responses are called response
times (RT), and are distinguished from
classical reaction time where special steps are
taken to see that a response is nonanticipatory
and always follows the stimulus (Woodworth,
1938). In our discrete tracking task a re-
sponse could follow a stimulus as in classical
reaction time studies, or it could be anticipa-
tory as might be expected from a time-
sensitive S who had acquired expectancy
states. A better understanding of expectancy
was hoped for by freely allowing, measuring,
and analyzing anticipatory responses. The
RTs were measured as the difference in milli-
seconds, between the onset of a stimulus and
the occurrence of a correct response to it.
Consistent with Poulton (1952), Adams and
Xhignesse (1960), and Adams and Chambers
(1962), a positive sign was assigned when the
response preceded the stimulus, and a negative
sign when it followed.

The basic analysis was conducted on an
RT score for an S, which was the algebraic
mean of all his response times to individual
stimuli of a particular set of stimulus events.
Each S had a Unisensory Visual RT, a
Unisensory Audio RT, and, for each audio
delay interval, a Bisensory Visual RT and a
Bisensory Audio RT score.

RESULTS

*Unisensory-bisensory audio compari-
sons.*—Figure 1 is a plot of group
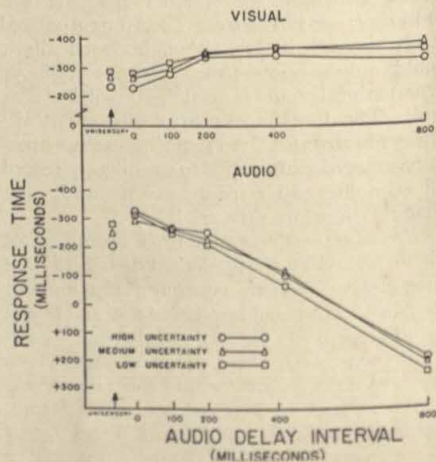


FIG. 1. Mean response times to unisensory
and bisensory visual and audio stimuli.

means for Unisensory and Bisensory RT scores for visual and audio. The lower part of Fig. 1 presents the data that are most critical for assessing the one-channel hypothesis. The hypothesis predicts that Bisensory Audio RT will be lengthened when the interstimulus interval is very brief because $S$ must have a finite period of time to process the visual stimulus and its response. Figure 1 shows the expected effect, and it is evident for all groups, particularly at the zero audio delay interval. Using the $t$ test for related measures, a comparison was made for each group between Unisensory Audio RT scores and Bisensory Audio RT scores at delay intervals of 0, 100, and 200 msec., where the effects of psychological refractory period should be most evident. For all three groups, the Bisensory Audio RT was significantly poorer ($P < .01$) than the Unisensory Audio RT at the zero interval, but only Group HU had significant retardation at the 100- ($P < .01$) and 200- ($.01 < P < .05$) msec. audio delays. Beyond the three initial points, the Bisensory Audio RTs had near zero or positive values and indicate the presence of anticipatory responding.

The greater persistence of significant decrement in the Bisensory Audio RT scores for Group HU is in line with the expectancy hypothesis. However, an analysis of variance of Unisensory Audio RT scores gave a significant $F$ ratio ($.01 < P < .05$), with Group HU having the lowest score, which could accentuate the differences. Examination of the unisensory audio data suggested that this was due to the differential presence of anticipatory responding among the groups. Because virtually all studies of psychological refractory period have used discrete reaction time tasks where special steps are taken to avoid

the influences of anticipation in performance measures, it is of special interest to see if the same support can be found for the expectancy hypothesis when only nonanticipatory measures are used. Using the same approach as related experiments that dealt with anticipation (Adams & Chambers, 1962; Adams & Xhignesse, 1960), three classifications of individual RT measures were made as a means of examining anticipatory behavior: Beneficially Anticipatory RT values in the range of $\pm 133$ msec. that had less off-target time than ideal RT values (Klemmer, 1956, 1957) and, by giving little or no off-target time, allow the reasonable inference that positive anticipatory mechanisms were operating; Nonanticipatory RT values which were less than $-133$ msec., and in the range for classical reaction time where $S$'s response occurs substantially after the stimulus; and Detrimentally Anticipatory RT values which were greater than $+133$ msec. where $S$ responded well ahead of the stimulus and could net as much, and often more, off-target time than if he had waited for the stimulus to occur and responded as in classical reaction time. An $S$'s RT score in each of these classifications was the algebraic mean of all his individual RTs in a classification for a particular experimental condition.

Groups LU, MU, and HU had 3, 4, and 11% of their individual unisensory audio RTs, respectively, that could be classified as either beneficially or detrimentally anticipatory. Using only Nonanticipatory Audio RTs, a Nonanticipatory Audio RT score was computed for unisensory audio and for bisensory audio under delay conditions of 0, 100, and 200 msec. A $t$ test for related measures was made between unisensory and bisensory audio under each of the three delay

conditions. None of the tests achieved the .05 level of significance for Group LU, and Group MU had a *t* ratio at the .05 level for the zero interval only. Group HU had a *t* ratio significant at better than the .01 level for the zero delay, and *t*'s significant at the .05 level for both the 100- and 200-msec. intervals. The same trend in support of the expectancy hypothesis is evident as before.

*Unisensory-bisensory visual comparisons.*—The upper part of Fig. 1 shows group mean RTs for unisensory and bisensory visual performance. An analysis of variance test revealed unisensory visual performance to be significantly poorer than unisensory audio ($P < .01$). While this is the traditional finding of visual RT being slower than audio RT, the two unisensory series are not directly comparable because the unisensory visual series had less time uncertainty than unisensory audio. The dominant trend in bisensory is for *S*s to temporarily withhold their visual response as a function of the audio delay interval, and this trend is about the same for all groups. As with audio performance, *t* tests for related measures were run between unisensory and bisensory visual RT scores. None

were significant at the zero delay interval but all were significant at 100 msec. and beyond ($P \le .05$).

*Characteristics of bisensory performance.*—To see if further insight could be obtained into reasons for the patterns of response decrement, a new measure was devised for the bisensory data, called the Difference RT, which gave the time between the two bisensory responses. This measure was used to check whether the amount and trend of decrement could be related to the way in which *S*s lagged the audio response behind the visual response as a function of uncertainty of the series. The expectancy hypothesis suggests that Group HU would lag the most because they would have developed higher expectancies for longer delay intervals. The Difference RT is defined by the formula (Bisensory Visual RT) − (Bisensory Audio RT) + (Audio Delay Interval), where the value for the delay interval is always positive, and the algebraic convention for RT is retained. The Difference RT is positive when the audio response follows the visual response, and negative when it precedes. Figure 2 shows the plot of mean Difference RTs. All individual RT values, plus and minus, entered the group means in Fig. 2. The three groups all have the same general trend, with Group HU at a higher level throughout (longer lags). Figure 2 shows that when the audio delay interval was zero or quite small, *S*s lagged the audio stimulus by a small amount. But as the audio delay interval increased, *S*s increased the lag of audio responses correspondingly. It would appear that the simultaneous occurrence of the visual and the audio stimulus is a cue for a rapid visual response and a small lag of the audio response. But if there is audio delay, and the visual stimulus
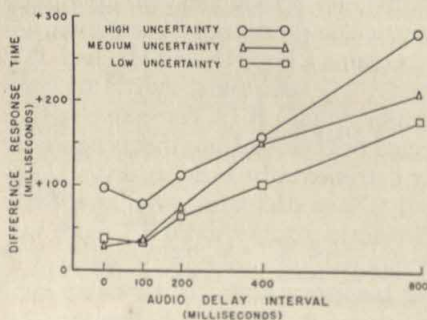


FIG. 2. Mean Difference RT values showing time lag between the visual and the audio response in bisensory tracking as a function of the audio delay interval.

TABLE 2

PERCENTAGES OF THE THREE POSSIBLE ORDERS OF BISENSORY RESPONSE PAIRS

| Order of Response Pair | Group and Audio Delay Interval (in Msec.) | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low Uncertainty (LU) | | | | | | Medium Uncertainty (MU) | | | | | | High Uncertainty (HU) | | | | | |
| | 0 | 100 | 200 | 400 | 800 | All Intervals | 0 | 100 | 200 | 400 | 800 | All Intervals | 0 | 100 | 200 | 400 | 800 | All Intervals |
| Visual first | 36 | 30 | 37 | 40 | 44 | 32 | 32 | 34 | 40 | 55 | 41 | 37 | 56 | 53 | 60 | 57 | 61 | 57 |
| Audio first | 25 | 15 | 14 | 20 | 16 | 16 | 28 | 18 | 24 | 16 | 20 | 20 | 11 | 13 | 12 | 12 | 10 | 12 |
| Simultaneous | 39 | 55 | 49 | 40 | 40 | 52 | 40 | 48 | 36 | 29 | 39 | 43 | 33 | 34 | 28 | 31 | 29 | 31 |

Note.—Percentages are based on the total number of correct response pairs made by a group under the conditions specified by a column heading.

is on momentarily by itself, then these delay states of affairs become a cue for delays in the visual response and longer lags of the audio response, although Fig. 1 shows that when the delay interval is 800 msec. the lag is poorly timed because Audio RTs are detrimentally anticipatory by 200–300 msec. Thus, S learns to interpret the immediate temporal properties of stimuli, and responds differentially to them, and Fig. 2 shows that the overall probability structure of delay intervals tends to differentially influence group mean performance in each case. The small intervals of 0, 100, and 200 msec. are most critical for issues in question, and an analysis of variance (Lindquist, 1953, Type I) was performed on the Difference RTs, with uncertainty a between-Ss variable and audio delay a within-Ss variable. Both main effects were significant at the .02 level. The same analysis was performed on Difference RTs computed from nonanticipatory responses, and again both main effects were significant ($P < .01$).

The Difference RTs show that Group HU lags the audio response more, which means longer RTs when audio delay intervals were small, and thus the decremental effect which is the evidence for the one-channel hypothesis. A related implication is that Group HU should have more bisensory response pairs occur in a visual-audio sequence rather than in a joint, simultaneous fashion. To evaluate this, the three possible sequences of response pairs were tabulated for each group, and the results are shown in Table 2. The order of responses can be either visual first, audio first, or simultaneous responding. The criterion for simultaneity was the two responses within ±33 msec. (Adams & Chambers, 1962). It is noteworthy that each audio delay interval has a percentage of response pairs in each of three orders, and there is no marked tendency towards "grouping," or simultaneous responding, to be particularly concentrated at very small intervals as Vince (1948) and Welford (1959) hypothesize would occur because of a stimulus pair being perceived as an entity under these conditions. Group differences are evident, particularly when the percentage is taken over all intervals, and the trends are ordered in accord with the expectancy view. Group HU had a dominant tendency to make the visual response first and lag the audio response. Group LU had stimulus pairs that almost always occurred with a very small time separation, and

they tended to respond with more simultaneity. Group MU had an intermediate position. To test these differences, each S was given a score in each of the three categories of response order that was the total number of responses made over all delay intervals. These were called the Visual First score, the Audio First score, and the Simultaneous score. A simple analysis of variance was performed for each of the three sets of scores, and the Visual First scores and the Simultaneous scores had between-groups differences that were significant at the .01 level. Audio First scores were significant between the .01 and .05 levels. The importance of Audio First pairings is not readily interpretable, but it could represent a tendency towards error in closely timed simultaneous responding.

## DISCUSSION

The results are consistent with the expectancy hypothesis. Decrement in response to the second of two closely spaced stimuli was greatest when the stimulus series had high time uncertainty, and was reliably less when time uncertainty was moderate or low. The reason was that Ss in the high uncertainty group had a greater likelihood of receiving a visual-audio stimulus sequence with a relatively long audio delay, and they learned to respond more frequently with a visual-audio order and to lag the audio response longer. The Ss had an expectation, or set, for the visual stimulus to come on first and for the audio stimulus ordinarily to be delayed and require a lagged response. Even when the two stimuli were presented simultaneously and Ss had direct and immediate information that a lag was not required, the audio response was still delayed. In fact, the Bisensory Audio RTs at the zero interval were the longest of any obtained. Thus, in this study, the decremental effect that has come to exemplify psychological refractory period

emerges as a learned tendency to respond with a visual-audio sequence and to lag the audio response when extensive practice has been given under conditions of temporal uncertainty.

The findings do not allow unequivocal rejection of the one-channel hypothesis, because even Group LU with low temporal uncertainty in their stimulus series had a significant amount of decrement in Bisensory Audio RT scores for stimulus pairs that occurred simultaneously. Remembering the wealth of temporal anticipation effects present in the data, this finding could be interpreted to mean that there is a one-channel decision mechanism that momentarily delays each S-R sequence, but that it cannot find a meaningful place in any scientific account of human behavior that does not give central focus to temporal expectancy states. Nevertheless, when these data are weighed with those of the Adams and Chambers (1962) study, there are defensible grounds for questioning the one-channel view. Using the same task, Adams and Chambers presented findings that dovetail with those here because one of their experimental conditions was complete temporal and event certainty. Not only did they find impairment absent, but they found that time and event certainty were the conditions for bisensory performance actually being superior to the aggregate unisensory control performances because of the influences of temporal anticipation. The conclusion from these two investigations is that given sufficient temporal certainty of events, S can process at least two simultaneous stimulus series with the same proficiency as a single stimulus series providing there is event certainty. It remains to be fully determined, however, whether a necessary central delay time exists for the resolution of *event uncertainty*. Adams and Chambers found first evidence for impairment under conditions of event uncertainty and temporal certainty.

## SUMMARY

An experiment was performed to test the hypothesis of psychological refractory period

that is offered to account for the established finding that response to the second of two closely spaced stimuli shows decrement. One line of explanation argues for a central decision time, where time must be allowed for processing the first stimulus and response before the second sequence can be undertaken. A competing explanation is the expectancy hypothesis which ascribes decrement to $S$'s past experience with the random array of interstimulus intervals that is usually used in experiments on this topic. Through practice, $S$ comes to expect a longer delay and the decrement is because he is not optimally ready to respond.

The experiment involved a two-dimensional, bisensory discrete tracking task. The statistical structure of interstimulus time intervals was the experimental variable aimed towards discriminating between the two hypotheses by asking if decrement could be a function of the temporal organization of stimuli. The results supported the expectancy hypothesis. Reliably less decrement was found for $S$s who trained on a stimulus series with a predominance of small time intervals and could learn behavior appropriate to them.

## REFERENCES

ADAMS, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58**, 55–79.

ADAMS, J. A., & CHAMBERS, R. W. Response to simultaneous stimulation of two sense modalities. *J. exp. Psychol.*, 1962, **63**, 198–206.

ADAMS, J. A., & XHIGNESSE, L. V. Some determinants of two-dimensional visual tracking behavior. *J. exp. Psychol.*, 1960, **60**, 391–403.

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Holt, 1959.

CRAIK, K. W. J. Theory of the human operator in control systems: II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, **38**, 142–148.

DAVIS, R. The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, 1956, **8**, 24–38.

DAVIS, R. The human operator as a single channel information system. *Quart. J. exp. Psychol.*, 1957, **9**, 119–129.

DAVIS, R. The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, 1959, **11**, 211–220.

ELITHORN, A., & LAWRENCE, C. Central inhibition: Some refractory observations. *Quart. J. exp. Psychol.*, 1955, **7**, 116–127.

HICK, W. E. The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1**, 36–44.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1957, **54**, 195–200.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MARILL, T. The psychological refractory phase. *Brit. J. Psychol.*, 1957, **48**, 93–97.

POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99–112.

POULTON, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229.

TELFORD, C. W. Refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–35.

VINCE, M. A. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, **38**, 149–157.

WELFORD, A. T. The "psychological refractory period" and the timing of high-speed performance: A review and a theory. *Brit. J. Psychol.*, 1952, **43**, 2–19.

WELFORD, A. T. Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, 1959, **11**, 193–210.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

# THE COURSE OF EMOTIONALITY IN THE DEVELOPMENT OF AVOIDANCE [1]

HOWARD S. HOFFMAN and MORTON FLESHLER

*Pennsylvania State University*

Dual process theories of avoidance (Hull, 1943; Miller, 1951; Mowrer, 1950; Solomon & Wynn, 1953) are predicated on the assumption that the acquisition of the instrumental avoidance response involves, and moreover depends upon, the concurrent acquisition of a conditioned emotional response (sometimes identified as fear or anxiety). The purpose of the present investigation was to assess the conditioned emotional responses (CERs) that occur during the acquisition of avoidance.

The results of such a study provide a test of dual process theory because the failure to demonstrate a meaningful relationship between emotional and instrumental behavior would cast considerable doubt on the dual process position. To the extent that meaningful relationships are observed, however, the results would provide an opportunity to evaluate the sequence of changing interactions between instrumental and emotional behaviors which, according to dual process theory, occur during acquisition. Theoretical specification of this sequence of interactions has, of necessity, been somewhat speculative. Experiments which have documented the acquisition of avoidance behavior have either inferred the state of concurrent emotionality from crude indices (bolus counts, incidence of

freezing, trembling, etc.) or more frequently from the effects of various experimental operations upon the avoidance response itself (Solomon & Brush, 1956). Studies which have focused upon the development of emotional responses, on the other hand, have seldom employed avoidance procedures. One exception, however, is a study by Black (1956) which assessed cardiac reaction during the acquisition of avoidance. Although Black's results appear to offer some support for dual process theory, they are difficult to interpret because he also found that about two-thirds of the total cardiac reaction could be attributed to the muscular occurrence of the instrumental response itself.

The approach of the present study was to track the course of emotionality by developing an avoidance response while Ss were engaged in positively reinforced ongoing behavior. In this paradigm, the index of emotionality is the decrement in rate of positively reinforced responding (conditioned suppression) which occurs during the presentation of a warning signal which precedes electrical shock. Conditioned suppression has been examined extensively and found to provide a sensitive and reliable index of emotional responses (Brady & Hunt, 1955; Estes & Skinner, 1941). The present study differs from the usual conditioned suppression experiment in one major respect. In studies of conditioned suppression, the noxious event is unavoidable. In the present arrangement, on the other hand, a specific instrumental response, during

the CS, terminates that stimulus and prevents the noxious event.

## METHOD

*Subjects.*—The Ss were 12 female rats of Sprague-Dawley stock. They were approximately 10 mo. old at the start of the experiment.

*Apparatus.*—The experimental chamber was a sound insulated Skinner box fitted with a one-way vision observation port. In the middle of the front wall was a recess to which Noyes food pellets were delivered. On either side of the recess was a manipulandum. To the right was a bar of 0.25 × 0.75 in. aluminum which projected 1.5 in. into the test chamber at a height of 1.5 in. and which actuated a microswitch when a downward force of 20 gm. was applied. At the left, 1.5 in. above the floor was a plate (1.5 in. square) which protruded 0.5 in. from, and was parallel to, the front wall. When a horizontal force of at least 15 gm. was applied perpendicular to the plate, it actuated a second microswitch.

The walls, manipulanda, and grid floor were wired to carry electrical shock, and during shock the polarity of the grid bars, the manipulanda, and the walls was continuously scrambled, so as to make unauthorized escape highly improbable. The shock power was supplied by an Applegate constant current stimulator set at 1.5 ma.

Acoustic signals were delivered through a 5-in. speaker mounted on the back wall of the chamber. Out of a second speaker, mounted at the side, white noise was continuously presented in order to mask sounds having external origin. The warning signal was a pure tone at 3500 cps with an intensity of 88 db. re .0002 dynes per cm.², when measured in front of the speaker. The tone was generated by a Hewlett-Packard audio oscillator.

A series of timers, steppers, and relays was used to establish the several stimulus-response contingencies which the research demanded. The circuitry was such that a response (either bar press or plate press) was defined as the initial closure of the corresponding microswitch. Thus, holding responses had no effect on the program.

All stimuli and responses were recorded on an Esterline-Angus operations recorder. In addition, counters were used to record the number of responses occurring during the tone and during the periods which preceded and followed each tone. A Standard Electric timer was used to measure the latency of the bar press response to the tone.

*Procedure.*—All Ss were treated alike in each of the several stages of training. The rats were first taught to escape from shock. After 20 min. of adaptation to the box, shock was turned on periodically; only a bar press terminated shock. After 50 presentations of shock, the median escape latency had stabilized at 0.75 sec. Twenty-five additional shocks were then delivered on each of two successive sessions in order to establish the bar press to terminate shock, as a well-learned habit. During these two sessions, the median escape latencies were 0.74 and 0.75 sec., respectively. The Ss were then placed on restricted feeding and from then on were maintained at 75% of their previously determined free feeding body weights. During the period of weight reduction Ss were placed in the box with the bar removed, so that generalized emotional responses could extinguish. In successive sessions, Ss were trained to the food magazine, shaped to press the plate and were run on a six-response fixed ratio schedule of reinforcement. (The Ss were exposed to the fixed ratio schedule in order to produce efficient response topographies.) A variable interval schedule of reinforcement with a mean of 30 sec. was in effect for the remainder of the experiment. Twenty sessions of plate pressing for food on the VI schedule were given to permit response rates to stabilize. By the end of these sessions, the rates had leveled off at a median value of 42 responses per min. These and all subsequent sessions each lasted 2.5 hr. and were run every other day.

During the next two sessions, while Ss pressed the plate for food, 20 tones, each lasting 60 sec., were presented without shock at intervals of 5 to 7 min. This was done so that in later phases of the study, it would be possible to determine whether or not observed decrements in plate pressing during tone were attributable to the pairing of tone and shock.

The procedure was the same for the next three sessions except that the bar had been replaced. In these sessions a bar press which occurred during a tone terminated that tone. This procedural detail was initiated in order to determine whether or not, prior to its pairing with shock, the offset of tone would reinforce the bar press.

In the following session, avoidance training was initiated. The tone was programed to remain on for 70 sec. with shock programed to occur during its final 10 sec. A bar press at any time during the tone, but prior to
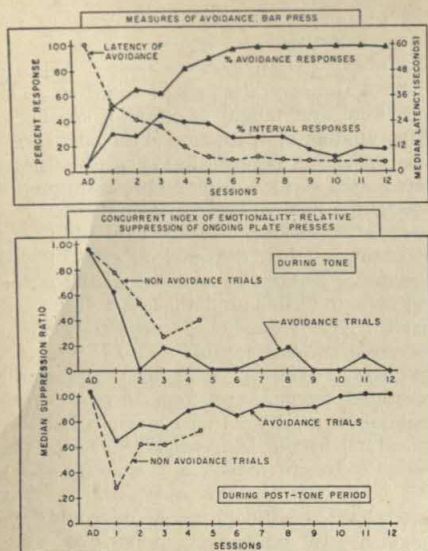
FIG. 1. Measures of avoidance and concurrent indices of emotionality throughout the course of acquisition. (AD refers to the final session of tone adaptation.)

shock onset, terminated that tone and prevented the occurrence of the shock. A bar press during shock terminated both the tone and the shock. Since the program of VI food reinforcement for plate presses was independent of the sequence of tone-shock pairings, food reinforcement could occur at any time during either the tone or the shock. Avoidance training was terminated after 12 sessions (20 tones per session).

*Experimental measures.*—For each $S$ on each trial, the following information was obtained: (*a*) whether or not shock was avoided; (*b*) whether or not a bar press occurred in the 60-sec. interval that ended with the onset of tone (hereafter this interval will be identified as the pretone period); (*c*) the latency of the bar press during tone; (*d*) the rate of plate press during the 60-sec. pretone period, i.e., the number of plate presses during the pretone period divided by 60; (*e*) the rate of plate press during the tone, i.e., the number of plate presses during the tone divided by the duration of the tone (as determined by the latency of the bar press); and (*f*) the rate of plate press during the 60-sec. posttone period that began with offset of tone.

On each trial and for each $S$, Measures $d$ and $e$ and Measures $d$ and $f$ were then combined to form two suppression ratios. The first suppression ratio was the rate of plate press during tone divided by the rate of plate press during the pretone period, and serves as an index of the relative emotionality during the tone. The second suppression ratio was the rate of plate press during the posttone period divided by the rate of plate press during the pretone period, and serves as an index of the relative emotionality during the posttone period. These ratios were the entries employed in all subsequent statistical analyses of concurrent emotionality.

## RESULTS

Figure 1 shows the session-by-session growth in avoidance, and also the changes in relative emotionality which develop concurrently. The top section of this figure shows the percentage of shocks that were avoided on each session, the percentage of times one or more bar presses occurred in the pretone period, and the median latency of the bar press during tone.

As seen in Fig. 1, during the final session of tone adaptation, ("AD" in the figure) $S$s seldom pressed the bar during either the tone or the pretone period. Since three sessions of tone adaptation had preceded this one, it is clear that, prior to its pairing with shock, the tone exhibited little, if any, control over the bar press. This result suggests that the tone itself was not intrinsically aversive, and that tone offset was not intrinsically reinforcing. Avoidance conditioning was instituted on Session 1. From Session 1 on, if the bar was pressed during the 60-sec. warning period, it terminated the tone and permitted $S$ to avoid shock. If, however, a bar press during tone had a latency of more than 60 sec., it also occurred during shock and hence represented an escape response. Although both the shock and the tone were programed to terminate after 10 sec., no $S$ ever permitted shock to remain on for longer than 2 sec. During the first session of avoidance conditioning the

median latency of escape was 0.79 sec. and the latency did not systematically change as sessions progressed. This result is in no way surprising since Ss had previously received extended training on escape.

As seen in Fig. 1, the tendency to avoid increased with each session until by Session 6, 99% of the shocks were avoided. The tendency to bar press in the pretone period (interval responses) increased during the first three sessions, but with continued training gradually declined. The latency of the avoidance response simultaneously decreased, until by the end of Session 6, it had stabilized at about 6 sec.

A series of $t$ tests for related measures was conducted on the frequencies of interval vs. avoidance responses during each of the first four sessions. The values of $t$ ($df = 11$ for each) were 2.87, 4.01, 2.43, and 3.72 for Sessions 1 through 4, respectively. Since each of these values is significant at the .05 level (for a two-tailed test), it is clear that the three functions which appear in the top section of Fig. 1 represent the development of a well-discriminated avoidance behavior.

The bottom sections of Fig. 1 show the several indices of emotionality that were obtained on each session during the acquisition. The solid line in the middle section of Fig. 1 shows the median suppression ratio during tone for trials on which an avoidance response occurred. It can be seen that the tone caused essentially no suppression during the final adaptation session but that with the introduction of shock, it rapidly developed the capacity to suppress ongoing plate presses.

The dashed line in the middle section of Fig. 1 shows the median suppression ratio, during tone, on those trials during which $S$ failed to avoid.

Since, as sessions progressed, the number of these nonavoidance trials decreased rapidly, the data for Sessions 4 and 5 have been combined. No data are shown beyond Session 5 because, in this period, the number of nonavoidance trials was too small to yield reliable measures of suppression.

The suppression ratios obtained on avoidance trials provide an initial test of the dual process position, since a theoretical interpretation of the discriminated avoidance which occurred from Session 1 on (Fig. 1) must assume that at least during acquisition, a CER to the warning stimulus also occurred.

If, on a given trial, there were no systematic differences between the rate of plate press in the pretone period and the rate during tone, then during a given session, the number of suppression ratios above one should equal the number below one, i.e., the median suppression ratio would be one. If, as predicted by theory, the tone consistently evoked a CER, the median suppression ratio should be less than one. Sign tests conducted on the suppression ratios during tone, on avoidance trials, provided support for a dual process interpretation since each value of $z$, for Sessions 1 through 12, was greater than 4.0 ($P < .01$ for a two-tailed test, in each case).

There is, however, a question of the degree to which the low value of these ratios reflects the cessation in positively reinforced responding which must occur when $S$ leaves the plate and executes the bar press. Two sources of information were used to assess this question. First, a series of sign tests were conducted on the suppression ratios, during tone, for nonavoidance trials. In Sessions 1, 2, and 3, as well as in Sessions 4 and 5 combined, the values of $z$ were all greater than 2.82 ($P < .01$ in each

case). Thus, even on trials which were unconfounded by the occurrence of the avoidance response, the tone generated a substantial degree of suppression. Secondly, Ss were observed throughout acquisition. In general, behavior in the presence of the tone appeared to involve considerable emotionality. Such plate presses as occurred, were performed in a tentative manner and the instrumental avoidance response involved an extremely slow sequence of movements, even on trials with latency as short as 6 sec. Both behaviors stood in sharp contrast to the quick energetic movements which typified the plate press during the pretone period and the instrumental bar press during shock. Finally, it may be noted that even if the suppression ratios on avoidance trials were adjusted for the maximum time *necessary* to execute the bar press (approximately .79 sec. as estimated from the latency of the bar press during shock), only a small increase would occur and the general configuration of the data would be unaltered.

A second assertion derived from dual process theory, is that during acquisition the tendency to avoid will be directly related to the CER magnitude. As seen in Fig. 1, suppression, during tone, was consistently greater on avoidance trials than on nonavoidance trials.

A Mann-Whitney test was used to assess the reliability of these data. For Sessions 1 and 2, the differences seen in the middle section of Fig. 2 were highly significant: $z = 3.15$ in Session 1 and $z = 4.02$ in Session 2 ($P < .01$ in both cases). The values of $z$ in Session 3 and in Sessions 4 and 5 combined, while in the direction predicted by theory, only tended toward significance at the .05 level ($z = 1.79$ in Session 3 and $z = 1.92$ in Sessions 4 and 5 combined).

Despite the failure to attain significance for the later sessions, the general pattern of these results supports the proposition that early in acquisition, the probability of the avoidance response is directly related to the suppressing capacities of the tone.

The bottom section of Fig. 1 shows the course of emotionality in the posttone period on avoidance trials and on nonavoidance trials. A third assertion of dual process theory is that reinforcement for successful avoidance consists of a reduction in conditioned emotionality. It can be seen that, during the first few sessions, even when avoidance responses occurred, positively reinforced behavior tended to be suppressed during the posttone period. This finding raises the question of whether successful avoidance responses were actually accompanied by a reduction in emotionality and if so, at what point did this effect begin? Sign tests were also used to assess this question. However, in those tests, the paired items were the two suppression ratios (tone and posttone) from a given animal on a given trial where shock was avoided. The tests revealed that in Session 1, suppression during the posttone period was not reliably different from suppression during the tone; $z = 1.85$ ($P > .05$). From Session 2 on, however, with the occurrence of an avoidance response, the level of relative suppression underwent a statistically significant decrease. Each of the values of $z$, for Sessions 2 through 12 was greater than 3.74 ($P < .01$ for a two-tailed test, in each case). Thus, if reinforcement consists of a reduction in emotionality, these data suggest that the avoidance response was reinforced only from the second session on.

Figure 2 shows the percentage of avoidance response per block of five

trials and serves to illustrate the changes in performance that occurred within sessions. The within-session changes in emotionality are not shown, because when based on samples of only five trials, the random fluctuations of the several indices were of such magnitude as to obscure any underlying trends.

As seen in Fig. 2, during the final session of tone adaptation there were no systematic changes in the tendency to press the bar during the tone. With the introduction of shock at the end of tone, however, a performance emerged in which, during the initial five sessions, the tendency to avoid increased markedly within each session and decreased (to a lesser extent) in the 48-hr. interval between sessions. After Session 5, the performance had reached a stage in which very few shocks were received and such changes as occurred within and between sessions were small and unreliable.

## DISCUSSION

Interpretation of these results must recognize that the suppression ratio is, at best, an index of relative emotionality. It reflects the magnitude of a *change* in emotionality, but it provides no information about the absolute level of emotionality just prior to the change (i.e., the emotionality during the pretone period). Although none of the experimental measures in the present study were geared to provide an accurate assessment of the absolute level of emotionality, the data on the rate of plate presses during the pretone period is relevant to this question. During the final session of tone adaptation, the median pretone rate was 46 responses per min. During Session 1, when shocks occurred frequently, the median rate fell to 25 responses per min. However, as sessions progressed (and shock frequency decreased) the median pretone rate increased with each session, until by Session 7, it had reached 44 responses per
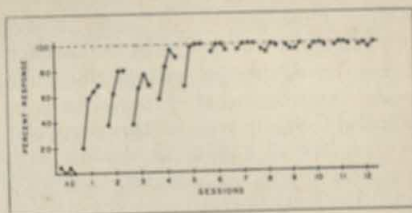


FIG. 2. Percentages of avoidance responses per block of five trials throughout the course of acquisition. (AD refers to the final session of tone adaptation.)

min. Thereafter, the median pretone rate remained relatively constant, never falling below 40 responses per min., nor rising above 50 responses per min. From these data, it may be hypothesized that with the introduction of shock, the absolute level of emotional reactivity increased greatly, but that as sessions progressed, it slowly declined. Apparently, the obtained suppression ratios (represented in Fig. 1) reflect changes in emotionality over and above a moving baseline of generalized emotional reactivity. In this respect, it should be noted that a number of investigators have shown that some form of generalized emotional reactivity is an important variable in conditioning (Spence, 1958; Spence, Farber, & Taylor, 1954). Moreover, the results of a previous study by Hoffman, Fleshler, and Chorney (1961) cast additional light on this issue. In that study, it was found that even after extensive training, certain rats would fail to avoid on the early trials of each session, but would achieve a high level of performance by the end of the session. The results of that study indicated that the occurrence of shock was the critical factor in this warm-up-like phenomenon and for this reason suggested that warm-up in avoidance reflects a motivational process. Apparently, as shocks occur, their emotional aftereffects persist and summate to produce a state of emotional reactivity which facilitates avoidance.

In the present study, warm-up was exhibited during acquisition (Fig. 2) and in the absence of evidence to the contrary, it is reasonable to assume that this

feature of the performance also reflects the action of the lingering emotional aftereffects of aversive stimulation. Although existing dual process theories have not formally treated this particular process, Spence (1956) in dealing with classical conditioning, suggests that ". . . the drive level operating at the time of the conditioned anticipatory response is a function of the residual effects of the internal response ($r_e$) to the noxious stimulus of the preceding trials. That is, such emotional responses are assumed to have a relatively persisting effect that extends well beyond the range of temporal intervals usually employed in conditioning experiments . . ." (p. 186). It is clear that a motivational process, such as Spence describes, can be readily incorporated within existing dual process theory and hence that the occurrence of warm-up is consistent with a dual process interpretation of avoidance.

## Summary

In order to examine the interplay between instrumental and emotional behavior during the acquisition of a discriminated avoidance response, rats were trained to press a bar to avoid shock while they were concurrently engaged in pressing a plate for food. The course of emotionality was tracked by assessing the several levels of suppression of ongoing plate presses during each of the various phases of the acquisition process. The results revealed a complex relationship between the level of performance on avoidance and the several concurrent indices of emotionality. In general, the results support the dual process hypothesis that conditioned emotionality controlled by the warning signal provides motivation for the avoidance response, while a decline in emotionality (with the offset of the signal) reinforces the response. The results also suggest that the lingering motivational aftereffects of aversive stimulation play an important role in the early phases of acquisition.

## References

Black, A. H. The extinction of avoidance responses under curare. Unpublished doctoral dissertation, Harvard University, 1956.

Brady, J. V., & Hunt, H. An experimental approach to the analysis of emotional behavior. *J. Psychol.*, 1955, 40, 313–325.

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, 29, 390–400.

Hoffman, H. S., Fleshler, M., & Chorney, H. Discriminated bar-press avoidance. *J. exp. Anal. Behav.*, 1961, 4, 309–316.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

Solomon, R. L., & Brush, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956. Pp. 212–305.

Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychol. Monogr.*, 1953, 67(19, Whole No. 354).

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W. A theory of emotionally based drive ($D$) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, 13, 131–141.

Spence, K. W., Farber, I. E., & Taylor, J. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, 48, 404–408.

# ON THE RELATIONS AMONG SOME FACTORS THAT CONTRIBUTE TO ESTIMATES OF VERTICALITY

C. R. CURRAN AND H. L. LANE

*University of Michigan*

Recent research has identified some factors that effect nonveridical perception of verticality (Asch & Witkin, 1948a, 1948b; Wapner, Werner, & Chandler, 1951; Wapner, Werner, & Morant, 1951; Werner, Wapner, & Chandler, 1951; Witkin, 1949, 1950, 1952; Witkin & Asch, 1948a, 1948b). Through the experimental manipulation of visual context, body tilt and body support, among other variables, visual, proprioceptive, labyrinthine, and tactile cues have all been implicated as contributors to erroneous estimates of the upright. Further progress in the analysis of perception of the upright would seem to require at least the following three steps:

*Isolation of the sensory events involved.*—This could be achieved through experimental control, requiring manipulation of variables one at a time, or through partialing out the effects of these variables, requiring multidimensional experimental design, or by both techniques. Earlier studies have often confounded the effects of sensory events in several modalities when investigating the relation between a relatively complex experimental operation and perception of the upright. For example, the procedure of tilting *S* has been the most widely used experimental operation in research on perception of the upright, yet it may produce concurrent changes in visual, proprioceptive, labyrinthine, and tactile stimulation. Werner, Wapner, and Chandler (1951) have interpreted the effect of body tilt on estimates of the upright as evidence for their sensory-tonic field theory of perception, which suggests that the degree of muscular involvement plays an important role in

determining these judgments. This inference must remain tentative, however, until the cluster of sensory changes effected by body tilt is experimentally analyzed. A similar case may be made for analyzing the complex effects of changes in the visual field. Witkin and Asch (1948b) have established that "the effect of the visual field upon the perceived upright tends to be stronger and more consistent the more richly articulated the field" (p. 782). However, richness of articulation is undoubtedly a multidimensional affair.

*Use of independent variables measured on ratio scales.*—Recent studies of perception of the upright have typically employed dependent variables measured on ratio scales (e.g., the angle between a rod called vertical by *S* and the true vertical) and one or more experimental treatments defined by nominal or ordinal scales (e.g., "tilted standing" and "tilted sitting"). Clusters of variables, such as support, lend themselves to nominal scaling; fractionation of these clusters should point to stimulus variables defined in terms of physical dimensions and thus measured (typically) on ratio scales. When *both* the dependent and independent variables are measured on ratio scales the predictive power of the findings is greatly enhanced, because the ratio scale contains the interval scale within itself (as well as the ordinal and nominal scales) (Stevens, 1960).

*Quantitative analysis of the effects of variables and their interactions in multidimensional experimental design.*—If a set of variables affecting perception of the upright has been isolated, if each can be measured on a ratio or, at least, interval scale, and if several levels or values of each variable are incorporated in an experimental design, then it is
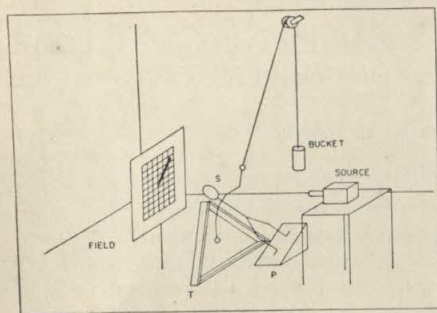
FIG. 1. Arrangement of apparatus for measurement of nonveridical perception of the upright.

possible to quantify their relative effects and their interaction effects in determining perception of the upright and to obtain the predictive power we desire.

The research to be reported represents a modest attempt to assess the effects of incorporating these three methodological improvements in an investigation of perception of the upright. The study describes the relation between changes in visual, proprioceptive, and labyrinthine stimulation and the degree of tilt of a rod, when it is reported to be vertical by the observer.

## METHOD

The major experimental variables were: the illumination of the visual field ($10^{-2}$, $10^{-3}$, or $10^{-4}$ ft-c), the degree of body tilt ($10°$ or $30°$), and the counterbalancing weight (equilibrium, [W], W plus 6 lb. and W plus 12 lb.). Parametric variables were: direction of tilt (left or right) and starting position of the rod (left or right).

Twelve male, naive undergraduates were $S$s in sessions lasting from 60 to 90 min. All $S$s had normal, uncorrected vision. Their heights ranged from 5 ft. 7 in. to 6 ft. 2 in., their weights from 131 to 179 lb. Earlier research has reported wide individual differences in judgments of the upright under distorting conditions (Witkin, 1949) but considerable consistency of judgment within individuals. A four-way experimental design was therefore employed, comprised of the three major variables (illumination, tilt, and

counterbalancing weight) and a "blocking variable" ($S$s). Direction of tilt and starting position of the rod were confounded with $S$s, so that the four levels of the blocking variable were: right tilt, right rod ($N = 3$), left tilt, right rod ($N = 3$), etc. Each $S$ gave 3 judgments under each of 18 combinations of the levels of illumination, tilt, and weight, presented in counterbalanced order.

Figure 1 is a schematic representation of the apparatus. A triangle (T) with base angles $60°$ and $80°$, was constructed from $2 \times 4$ in. boards and served as a tilt reference for $E$. The $E$ aligned the median plane of $S$ with the appropriate arm of the tilt reference by visual inspection before each trial. With $S$ standing erect, the light source, 4 ft. behind him, and the center of the stimulus field, 4 ft. in front, were at eye level and in approximately the same vertical plane. The $S$ stood on a pedestal (P) that was held firmly in place on the floor. The surface of the pedestal slanted upward at an angle that was set equal to the angle of tilt, so that $S$ was perpendicular to the surface when aligned with the tilt reference. When $S$ was tilted, he grasped the supporting rope in the hand contralateral to the direction of tilt. The rope passed over a pulley suspended from the ceiling (18 ft. high) and terminated in a bucket of cement, selected by $E$.

This cement weight was selected in the following way. At $10°$ tilt, a weight, $W_{10}$, was determined that was within .5 lb. of $3\%$ of $S$'s weight. At $30°$ tilt, $W_{30} = 25\%$ of $S$'s weight, $\pm.5$ lb. Under either tilt condition, the levels of the counterbalancing variable were then W, W + 6 lb., and W + 12 lb. The value of W under each condition of tilt approximated the weight that would just balance $S$. The moments of force around the pedestal are approximated by: $M (h_c \text{ sine } T) = Wh_a$ where M = weight of $S$, $h_c$ = distance from the fulcrum to $S$'s center of gravity, T = angle of tilt from the vertical, W = counterbalancing weight, and $h_a$ = distance from the fulcrum to $S$'s arm.

Under all experimental conditions, $S$ viewed the stimulus field monocularly, through a reduction tube (3 in. long, 2 in. in diameter) attached to a pair of goggles. In this manner, the visual field of the eye contralateral to the direction of tilt was restricted to the stimulus field, while vision in the homolateral eye was blocked.

The field had an illuminance of approximately $10^{-2}$ ft-c with unfiltered illumination from a source whose color temperature was approximately $1800°$K (34 v. applied to a

500-w. lantern slide projector). Wratten neutral density filters were inserted to produce illuminances of $10^{-3}$ and $10^{-4}$ ft-c giving the three levels of the illumination variable. The stimulus field consisted of a piece of white poster board, 60 × 40 in., with a luminous reflectance of approximately 0.83. A piece of linear graph paper, 8 × 8 in., with blue lines heavy-ruled at 1-in. intervals and light-ruled at .1-in. intervals, was mounted on the front of the poster board and aligned with the true vertical by means of a plumb line. (The luminous reflectance of the graph paper was approximately 0.77; at an illuminance of $10^{-4}$ ft-c no $S$ reported seeing the rulings while at $10^{-2}$ ft-c the paper and its rulings were visible to all $S$s.) Behind the poster board, a small motor rotated a shaft that punctured the board at its center and protruded $\frac{1}{2}$ in. A brass rod (reflectance 0.40) 12 in. long, $\frac{1}{8}$ in. diameter, was mounted at right angles to this shaft in front of the stimulus field for observation by $S$, and a 26-in. rod was mounted in parallel behind the poster board for observation by $E$. Twenty-four inches above the pivotal point of the rod at the center of the field, $E$ could read the point of intersection of the rod with a horizontal line, ticked at $\frac{1}{4}$-in. intervals; he could therefore read the position of the rod to an accuracy of about 0.5°. An ac motor with reduction gears swept the rod from its starting position, 50° ± 10° to the left or to the right of the vertical, toward the upright at a constant rate of 1.5° per sec. When $S$, who was observing the position of the rod under a given condition of tilt, weight, and illumination, judged it to be vertical, he said "stop" aloud, at which time $E$ pressed a normally closed switch to stop the motor. He then read the
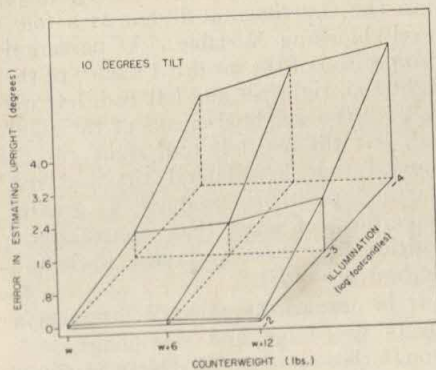


FIG. 2. The effect of counterbalancing weight, illumination, and tilt (10°) on perception of the upright.
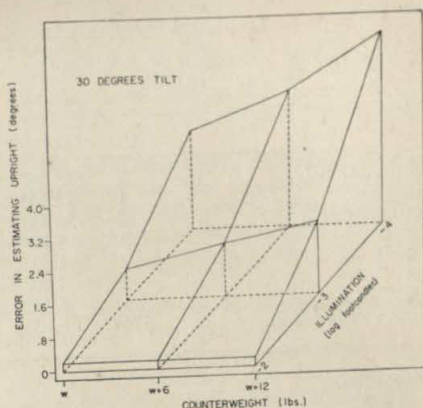


FIG. 3. The effect of counterbalancing weight, illumination, and tilt (30°) on perception of the upright.

rotary position of the rod, using a pencil light to read the horizontal scale. Following a trial, $S$ was instructed to close both eyes while $E$ changed tilt, weight, illumination, and rod position, in accordance with the protocol.

## RESULTS AND DISCUSSION

Figures 2 and 3 summarize the findings of this experiment by depicting the effect of the weight and illumination variables on estimates of the upright at each of the two levels of the tilt variable. Examination of these figures reveals the relative contribution of these three variables as well as their interaction effects. Clearly, the major variable is the level of illumination: the error solid grows rapidly in volume as the level of illumination is decreased at 10° tilt and even more rapidly at 30° tilt; this comparison reveals an Illumination × Tilt interaction. At relatively high levels of illumination the counterbalancing weight has little or no effect at 10° tilt and only a slight effect at 30° tilt. As illumination is decreased, however, the counterbalancing weight plays an increasing role in determining non-veridical perception of the upright. Comparison of Fig. 2 and 3 reveals a Weight × Tilt as well as a Weight

TABLE 1

ANALYSIS OF VARIANCE OF ESTIMATES
OF THE UPRIGHT

| Source | df | F |
|---|---|---|
| Illumination (I) | 2 | 525.26** |
| Tilt (T) | 1 | 72.15** |
| Counterweight (C) | 2 | 56.74** |
| I × C | 4 | 5.23** |
| C × T | 2 | 4.52* |
| I × T | 2 | 3.20* |
| I × C ×T | 4 | 4.33** |
| Within cells (MS) | 576 | (0.85) |

* P < .05.
** P < .01.

× Illumination interaction, and a Weight × Tilt × Illumination interaction. The solid obtained at 30° tilt is appreciably larger in all cells than that obtained at 10° tilt, showing the net effect of the tilt variable. As just indicated, there are also obvious Tilt × Weight and Tilt × Illumination interactions.

Table 1 presents an analysis of variance of the estimates of verticality. As anticipated, the variance attributable to replications within Ss is relatively small. All three main effects and their first- and second-order interactions are significant at the .05 level or beyond.

The present findings support Witkin's (1949) analysis of the relative contribution of visual as opposed to somesthetic cues in the determination of perception of the upright. These findings show, furthermore, that relatively few visual cues to the vertical can yield extremely accurate estimates of the vertical, even under marginal visibility, with the body tilted and delicately poised. It appears that the visual field need not be "richly articulated" to permit accurate estimates of the upright under distorting conditions. However, removal of these few visual cues by approximately halving the brightness of the field yielded almost a hundredfold increase in error in perceiving the upright.

When visual cues were minimized, at the lowest illuminance, and proprioceptive and cutaneous cues were minimized in the condition of equal moments around the pedestal, the degree of tilt was observed to have a considerable effect on the magnitude of the error in perceiving the upright. Inference from these findings suggests that, in the absence of visual and somesthetic cues to the upright, other cues to the static position of the body are available; perhaps the utricular otoliths, which are thought to play a role in static positional adjustments of the body (Geldard, 1953, p. 262) are the source of this stimulation.

When the condition of equilibrium is displaced through the addition of counterbalancing weights, greater effort is required on the part of S to maintain his balance and his alignment with the tilt reference. As the counterbalancing weight is increased, the magnitude of the error in judging the vertical is increased. This finding may be related to the qualitative prediction of Werner, Wapner, and Chandler (1951): "Within the framework of the sensory-tonic field theory of perception, the degree of muscular involvement is expected to be an important variable" (p. 346). The results of the present study also confirm the observation of these authors that the apparent vertical is shifted to the side opposite the direction of the body tilt. It will be remembered that the direction of tilt was confounded with the starting position of the rod and with Ss and incorporated into the experimental design as a four-level blocking variable. A posteriori comparison of the weighted means of the right rod, right tilt and left rod, left tilt Ss with the weighted means of the right rod, left tilt and left rod, right tilt Ss showed that homolateral tilt and rod starting positions produce a greater magnitude of error than contralateral positions of these variables (Scheffé's method; α = .05).

It is noteworthy that all interaction effects are large and significant. Although these interaction effects have not been demonstrated previously in the analysis of perception of the upright, this seems a plausible way for a perceptual

process to be controlled by the relevant variables. As indicated earlier, this type of analysis is facilitated through the use of multidimensional experiments with unidimensional variables sampled at several levels and measured on ratio scales.

## SUMMARY

Several variables that have been shown to influence the perception of the upright were incorporated in a multidimensional design to permit analysis of their several effects and interactions. Minimal visual cues had a dramatic effect in reducing nonveridical perception of the vertical. Distortion of body tilt and balance produced effects of lesser magnitude. All the first- and second-order interactions of these variables had large and significant effects on perception of the upright.

## REFERENCES

ASCH, S. E., & WITKIN, H. A. Studies in space orientation: I. Perception of the upright with displaced visual field. *J. exp. Psychol.*, 1948, **38**, 325–337. (a)

ASCH, S. E., & WITKIN, H. A. Studies in space orientation: II. Perception of the upright with displaced visual fields and body tilted. *J. exp. Psychol.*, 1948, **38**, 455–477. (b)

GELDARD, F. A. *The human senses.* New York: Wiley, 1953.

STEVENS, S. S. Ratio scales, partition scales, and confusion scales. In H. Gulliksen and S. Messick, (Eds.), *Psychological scaling: Theory and application.* New York: Wiley, 1960.

WAPNER, S., WERNER, H., & CHANDLER, K. Experiments on the sensory-tonic field theory of perception: I. Effect of extraneous stimulation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 341–345.

WAPNER, S., WERNER, H., & MORANT, R. B. Experiments on the sensory-tonic field theory of perception: III. Effect of body rotation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 351–357.

WERNER, H., WAPNER, S., & CHANDLER, K. Experiments on the sensory-tonic field theory of perception: II. Effect of supported and unsupported tilt of the body on visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 346–350.

WITKIN, H. A. Perception of body position and of the position of the visual field. *Psychol. Monogr.*, 1949, **63**(7, Whole No. 302).

WITKIN, H. A. Perception of the upright when the direction of the force acting on the body is changed. *J. exp. Psychol.*, 1950, **40**, 93–106.

WITKIN, H. A. Further studies of perception of the upright when the direction of the force acting on the body is changed. *J. exp. Psychol.*, 1952, **43**, 9–20.

WITKIN, H. A., & ASCH, S. E. Studies in space orientation: III. Perception of the upright in the absence of a visual field. *J. exp. Psychol.*, 1948, **38**, 603–614. (a)

WITKIN, H. A., & ASCH, S. E. Studies in space orientation: IV. Further experiments on the perception of the upright with displaced visual fields. *J. exp. Psychol.*, 1948, **38**, 762–782. (b)

# THE PERSPECTIVE ILLUSION: PERCEIVED SIZE AND DISTANCE IN FIELDS VARYING IN SUGGESTED DEPTH, IN CHILDREN AND ADULTS [1]

JOACHIM F. WOHLWILL

*Clark University*

Since the Renaissance it has been generally known that perspective drawings can convey a strong impression of depth in two-dimensional stimulus fields; curiously, however, this phenomenon has received little systematic attention on the part of psychologists. To be sure, Gibson, in his treatment of space perception, and of pictorial perception generally, has repeatedly pointed to the important role of perspective (Gibson, 1950, 1954, 1960); yet, apart from a largely exploratory study by Smith, Smith, and Hubbard (1958), the actual effects of perspective on perception have not been experimentally investigated.

The present study is concerned with one aspect of this problem, viz. the extent to which stimulus fields constructed according to the principles of perspective geometry will affect judgments of size and length within such a field. Thus, one might expect the height of an object located at the

bottom (i.e., apparent front) of a perspective drawing to be perceived as smaller relative to a similar object located at the top (i.e., apparent rear) of the drawing. In fact, just such an essentially illusory effect is dramatically illustrated by Gibson (1950, p. 182), by means of a perspective drawing of a corridor in which several barrel-shaped objects are depicted: the rearmost barrel appears strikingly expanded in size in comparison with the front one. More systematic investigation of these effects, and especially of their variation as a function of the characteristics of the stimulus fields responsible for the suggestion of depth should provide a clearer picture of this illusion, its magnitude and its determinants. Such a study should furthermore be of direct relevance to Gibson's (1950) general theory of space perception, which emphasizes the information to depth contained in the gradients of texture-density, etc., present in any two-dimensional projection of a three-dimensional stimulus field, be the projection retinal, photographic, or in the form of a perspective drawing.

The aim of this study is thus to investigate experimentally the effects of perspective drawings on the perception of relative linear extent in the plane of the drawing, with reference both to the perceived size of objects in this plane and the perceived distance between points in the plane. The principal variable manipulated in

the study is the nature of the "information to depth" contained in the field: first, the greater the amount of this information, as expressed in the *density* of the texture of the field, the greater should be the distorting effect of perspective; second, the introduction of redundancy into the field, expressed in terms of the patterning of the elements of texture, so as to enhance linear perspective, should likewise increase the effect. The effect should be maximal, finally, in a field portraying directly the geometrical relationships involved in a perspective transformation.

An additional variable of considerable interest in this domain of perception is that of the age of Ss. There is considerable evidence that spatial relationships generally exert relatively little influence on the perception of young children (cf. Wohlwill, 1960), so that one might postulate that the distorting effect of perspective is absent, or at least rather small in magnitude in early childhood, and will increase with age. Indeed, Glasser
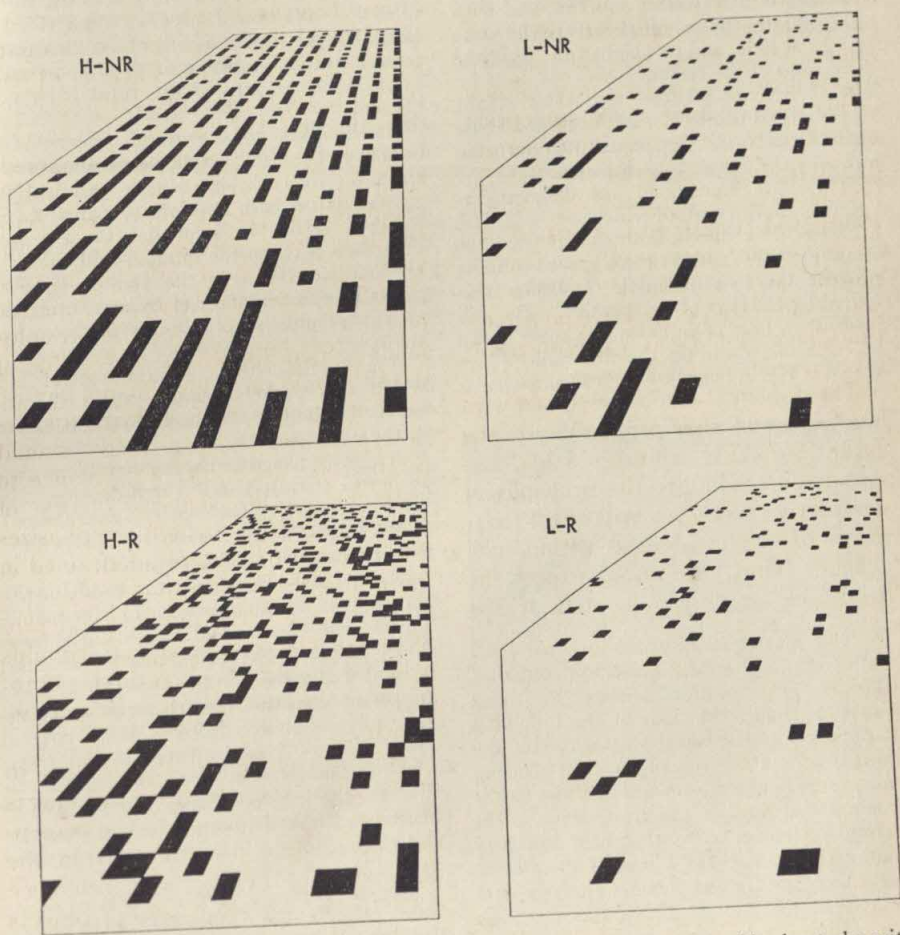


FIG. 1. Perspective drawings representing four of the stimulus fields utilized in the study, with texture-density (left vs. right figures) and randomness (top vs. bottom figures) as variables.

(1944) has claimed that young children rarely perceive a perspective illusion in a figure very similar to that of Gibson referred to above, and accordingly attributes the effect to the role of experience; however, no data are adduced in support of this statement, nor is the age of *S*s to which it is intended to apply further specified.

## METHOD

*Stimuli.*—Six different stimulus fields, drawn on sheets of drafting paper 23 × 29 in. in size, were employed in this study. The shape of all of these fields was uniform, consisting of a trapezoid superimposed on a rectangle; the bases of the trapezoid were 7⅝ in. and 12 in. and its height 7 in., while the dimensions of the rectangle were 12 × 18 in. The intersection of the two sides of the trapezoid at a point 12 in. above its lower base defined the vanishing point used for constructing the perspective drawings.

These fields were filled with different perspective drawings, all of which were based on a grid of 36 columns fanning out from the vanishing point and 62 rows spaced so as to produce the foreshortening of distance required by the laws of perspective. The grid itself, with lines of uniform thickness drawn in India ink, made up the first panel (illustrated in Fig. 2, below). Four of the remaining panels, shown in Fig. 1, were constructed by filling in selected cells from this grid in India ink, according to the following plan: (*a*) High density, nonrandom (Fig. 1, top left): Every third column was selected from the grid; within each the cells to be filled in were determined by a table of random numbers, so as to yield an average proportion of ⅔ of the cells filled in per column, or ⅑ of the total number of cells in the field. (*b*) Low density, nonrandom (Fig. 1, top right): From among those cells selected under *a*, a subset was chosen by means of a table of random numbers, consisting of 3/10 of the cells of *a*, or a density of .18 per column, i.e., .06 for the total field. (*c*) High density, random (Fig. 1, bottom left): From every column of the grid ⅕ of the cells were randomly selected to be filled in. (*d*) Low density, random (Fig. 1, bottom right): From among the cells selected for *c*, a subset consisting of 3/10 of these cells, i.e., .06 of the cells of the total field, was chosen.

It should be noted that the proportion of cells from the total grid that were filled in was the same for both high-density panels (.20), as well as for both low-density panels (.06). Further, it will be seen that the essential difference between the random and the nonrandom panels lies in the greater sense of linear perspective afforded by restricting the cells to a limited set of columns, in the case of the latter.

Finally, the sixth panel was a control field, which was entirely blank, except for the border of the field, drawn in in India ink as in the other panels.

These six panels will henceforth be referred to as: G (Grid); H-NR (high-density, nonrandom); L-NR (low-density, nonrandom); H-R (high-density, random); L-R (low-density, random); and C (control).

*Apparatus.*—The apparatus used for displaying the stimulus panels and for manipulating the size and distance variables is shown in Fig. 2. (Figure 2 also shows Panel G, as it was exposed in the apparatus.) The apparatus consisted essentially of a rectangular piece of plywood, 29.5 × 35 in., covered by a sheet of Plexiglas fitted into a frame which was attached by means of hinges to the side of the plywood base. The Plexiglas cover could thus be opened for the insertion and removal of the individual panels exposed behind it.

For the *distance* judgments, two points were chosen in the stimulus field of the panels, defining the distance to be bisected by *S*. The points were located on an imaginary line passing through the vanishing point of the
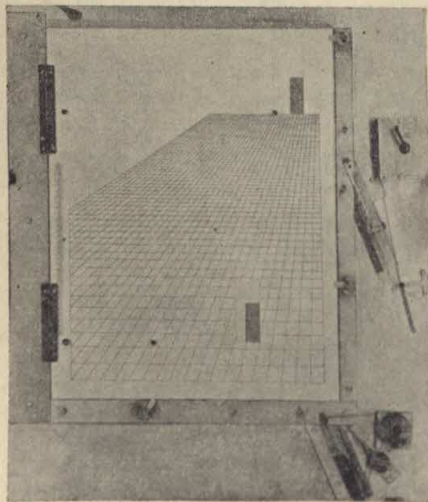


FIG. 2. Apparatus, with grid panel displayed.

perspective drawings, one point towards the top edge, the other near the bottom. Screws were driven through two corresponding spots on the Plexiglas cover, directly superimposed on the two points in the stimulus field as it was exposed underneath the cover.

An endless loop of fine nylon thread ran around a trapezoidal path marked by these two screws and two others near the left side of the Plexiglas cover and at the same heights as the former pair. This loop could be moved in either direction by pulling on a knot located along the vertical side, this movement causing a little red plasticine ball to travel up and down between the first two screws. A small indicator attached to the left vertical segment of the line, which moved along a ruler glued to the Plexiglas, enabled E to read off the height of the ball corresponding to the midpoint of the distance as perceived by S.

For the *size* judgments, two rectangular cutouts were made in each panel, through which the stimulus objects—two light-blue rectangular sheets of metal—appeared. One of the cutouts was at the top of the panel, the other towards the bottom, their lower-left vertices being located along a line through the vanishing point, well to the right of the line involved in the distance judgments (cf. Fig. 2).

The stimulus objects were attached to two wooden slides which could be moved up or down, along a line which was an extension of the diagonal of the rectangles, i.e., in an oblique direction relative to the field of the panel; movement of these slides thus caused the portion of the blue rectangle exposed through the cutout to vary in size, but with a constant ratio of height to width. Each slide was raised and lowered by turning a crank, to which it was connected by a set of pulleys.

This arrangement permitted continuous variation of both the top and bottom stimulus objects, so that either could be used as a variable and the other as a standard. The size of the standard rectangle, in terms of the length of the diagonal, was 7.5 cm.; the diagonal of the variable rectangle varied from 0 to 12 cm. In order to prevent Ss from utilizing the amount of white in the cutouts around the rectangles as a cue, the two cutouts were made unequal in size, the bottom one being 4.3 × 11.3 cm. (a size corresponding to the maximum size of the variable), whereas the top one measured 5.3 × 12.3 cm.

The whole apparatus was displayed in a vertical position to the S, by sliding it into a metal frame mounted vertically on a wooden base.

*Procedure.*—The adult Ss were tested in a room illuminated only by overhead fluorescent lighting. They were seated on a stool at a distance of 3 m. from the apparatus, so that their eyelevel was approximately even with the center of the stimulus panel. The S was told that the experiment concerned his ability to make judgments of size and distance relationships. Specifically, for the *distance* judgments he was told that E would make the little red dot travel upward along the line between the two screws, and instructed to say "stop" when he thought the dot was exactly midway between the screws. This wording of the instructions was intended to foster a set for objective, rather than phenomenal judgments; however, E discouraged S from attempting to "figure out" where the midpoint was intellectually. The S was further informed that he would have an opportunity to correct any setting if he was dissatisfied with it.

For the *size* judgments the instructions were similar, S being asked to say "stop" when the top (or bottom) rectangle was just equal in size to the bottom (or top) one.

In his manipulation of the red dot and of the variable rectangle, E always faced away from S, so as to avoid giving him any involuntary facial cues that might influence his judgment. Movement of the stimuli was carried out at a fairly even rate, although for the variable size stimulus perfect smoothness was not realizable with the apparatus as constructed.

Special procedures were used with the group of children from Grade 1, in order to ensure that they properly understood the instructions. These procedures consisted of a series of pretest judgments, involving, (a) marking the middle of two lines drawn on a 8½ × 11 in. sheet of paper, one line horizontal, the other oblique, (b) deciding when a bead which E moved along a string over the surface of a table, first parallel to the edge, then obliquely, was at the middle of the string, and (c) making a preliminary distance setting on the apparatus, without any panel underneath the Plexiglas cover. There were no instances in which any child gave an indication of failing to understand the instructions, either through questions directed to E, or through markedly deviant judgments on the pretest.

All of the children were routinely asked, "Is that it?" or "Is that where you want it?" after every judgment, and allowed to change it if they wished (as the adults had been also). This was done in an effort to minimize errors of anticipation. The children were further asked, after being presented with the

TABLE 1

MEAN DISTANCE SETTINGS (IN CM.)

| Groups | Stimulus Panels | | | | | | Combined | |
| | C | L-R | H-R | L-NR | H-NR | G | Mean | $s^a$ |
|---|---|---|---|---|---|---|---|---|
| Grade 1 | 21.91 | 22.19 | 22.45 | 22.23 | 22.97 | 23.12 | 22.48 | 1.35 |
| Grade 4 | 22.00 | 22.71 | 22.59 | 22.85 | 22.96 | 23.02 | 22.69 | 1.32 |
| Grade 8 | 21.40 | 22.01 | 22.00 | 22.27 | 22.44 | 22.72 | 22.14 | 1.15 |
| Adults | 21.06 | 21.08 | 21.20 | 21.19 | 21.43 | 21.80 | 21.30 | 0.86 |
| Combined Mean | 21.59 | 22.00 | 22.06 | 22.13 | 22.45 | 22.67 | | |
| $s^b$ | 1.42 | 1.81 | 1.78 | 1.82 | 1.87 | 1.77 | | |

Note.—Values tabled represent bisections of a 41-cm. vertical distance.
[a] SD of Ss' mean scores, based on between-Ss error terms for each age level.
[b] SD of scores for all Ss, based on residual error term calculated for each stimulus panel.

H-R and H-NR panels, whether they could think of anything that might look like what they were seeing.

*Design.*—For both distance and size judgments S made two judgments for each stimulus panel, one ascending, the other descending. Size and distance judgments were always made consecutively for any stimulus panel, before a new panel was exposed; half of the Ss always judged size first, the other half judged distance first. The six panels were presented in a Latin square design, in two different sequences, one being the reverse of the other. Regardless of the particular panels exposed, all Ss started with the ascending judgment on the first, third, and fifth panels and with the descending judgment on the second, fourth, and sixth panels. Finally, on the size judgments half of the Ss were tested with the top stimulus as the standard, and the other half with the bottom stimulus as the standard.

*Subjects.*—There were four groups of 24 Ss each, representing samples of children from Grades 1, 4, and 8 and college-age adults. The mean ages of these four groups in years and months were 7:1, 9:10, 14:0, and (approximately) 20:0, respectively. The school children all came from a lower-middle class grade school, and were thus not strictly comparable in IQ and related variables to the adults, who were college undergraduates (mostly freshmen and sophomores) enrolled in an introductory psychology course. All Ss reported they had normal eyesight, either uncorrected or corrected. (For the youngest children, the pretest with the apparatus given before the experiment proper allowed E to satisfy himself of the adequacy of S's eyesight.)

RESULTS

*Distance*

The distance between the two screws which was to be bisected was 47.3 cm. The vertical component of this distance (on which the recorded distance data are based) was 41.0 cm. Thus settings of the red dot larger than 20.5, the objective midpoint, would indicate an influence of suggested depth. This influence is apparent in the judgments at all ages and for all stimulus conditions, including Panel C, as shown by the means shown in Table 1. (Additional comparisons between the H and L panels and between the NR and R panels are provided in Table 2.)

TABLE 2

MEAN DIFFERENCES IN DISTANCE SETTINGS (IN CM.) BETWEEN PANELS VARYING IN DENSITY AND IN RANDOMNESS

| Groups | H vs. L | NR vs. R |
|---|---|---|
| Grade 1 | .50 | .28 |
| Grade 4 | .00 | .26 |
| Grade 8 | .08 | .35 |
| Adults | .18 | .17 |
| Combined | .19 | .26 |

Note.—Positive values indicate greater magnitude of illusion for first-listed panels.

As the analysis of variance of these data is, rather lengthy and complex, suffice it to present it in abbreviated form. It falls into four parts. The first, involving all between-$S$s effects (age, order, and their interaction), disclosed the variance due to age to be significant at better than the .01 level ($F = 5.95$, $df = 3/88$). The second, involving within-$S$s effects summed over direction (stimuli, singly and in interaction with age and order), showed the variance due to the stimulus fields to be highly significant ($F = 14.92$, $df = 5/440$), but no significant interactions. The third part, involving within-$S$s effects summed over stimulus panels (direction, singly and in interaction with age and order) showed a highly significant effect due to ascending vs. descending direction ($F = 152.54$, $df = 1/88$), as well as an interaction between direction and age significant at between the .05 and .01 levels ($F = 3.05$, $df = 3/88$). The fourth part, finally, is comprised of within-$S$s effects due to simple and higher-order interactions involving both the stimulus and the direction variable; here the simple interaction was significant at the .01 level ($F = 3.09$, $df = 5/440$).

These results may now be summarized as follows:

*Stimulus fields.*—The observed differences between stimulus fields are in good agreement with those postulated in our introduction: the effect of perspective was greatest for the G panel and least for the C panel (cf. Table 1) ; it was also greater for high- than for low-density fields, and greater for nonrandom than for random fields (cf. Table 2). Duncan's multiple range test (cf. Edwards, 1960, pp. 136 ff.) indicates the following comparisons between means (for all groups combined) to be significant at the .05 level: G vs. all others except H-NR; H-NR

vs. H-R, L-R, and C; L-NR vs. C; H-R vs. C; L-R vs. C. In addition, orthogonal comparisons (Edwards, 1960, pp. 140 ff.) between the two H vs. the two L fields, as well as between the two NR vs. the two R fields, both show differences significant at better than the .01 level.

*Age.*—The overall effect appears to *decrease* with age, except that Grade 4 $S$s showed slightly (but nonsignificantly) higher mean values than Grade 1 $S$s. An application of Duncan's multiple range test, however, shows that all three of the children's groups are significantly differentiated from the adults, but not from one another. Although there was a suggestion that the first graders were somewhat more influenced by the density variable than the other groups (cf. Table 2), the interaction between age and stimulus panels was not significant.

*Direction.*—A notable feature of the results was the finding of a very marked anticipation effect: for all age groups and stimuli combined, the mean ascending judgment was 21.44, while the mean descending judgment was 22.87. This effect itself interacted with age, being largest at the first- and fourth-grade levels, and considerably reduced at the eighth-grade and adult levels. There was also a significant interaction of this factor with the stimulus variable which was less consistent in nature.

*Order.*—The order-of-judgment variable (distance before vs. after size) failed to account for any significant portion of the variance, either singly or in interaction.

*Size*

The size of the standard rectangle (measured in terms of the length of the diagonal) was 7.5 cm. Thus settings of the variable smaller than 7.5,

## TABLE 3
### Mean Size Matches, in Cm. (POE = 7.5)

| Groups | Stimulus Panels | | | | | | Combined | |
|---|---|---|---|---|---|---|---|---|
| | C | L-R | H-R | L-NR | H-NR | G | Mean | $s^a$ |
| Grade 1 | 7.47 | 7.51 | 7.49 | 7.27 | 7.36 | 7.23 | 7.39 | 0.28 |
| Grade 4 | 7.39 | 7.49 | 7.53 | 7.35 | 7.33 | 7.36 | 7.41 | 0.29 |
| Grade 8 | 7.42 | 7.49 | 7.50 | 7.26 | 7.39 | 7.27 | 7.39 | 0.24 |
| Adults | 7.29 | 7.50 | 7.61 | 7.38 | 7.33 | 7.31 | 7.40 | 0.23 |
| Combined | 7.39 | 7.50 | 7.53 | 7.32 | 7.36 | 7.29 | | |
| $s^b$ | 0.32 | 0.34 | 0.39 | 0.35 | 0.37 | 0.41 | | |

Note.—Values tabled represent height at top of field judged equal to a 7.5-cm. height at bottom (see text for details).
 a See Footnote a, Table 1.
 b See Footnote b, Table 1.

when the standard was at the bottom of the field, and larger than 7.5, when the standard was at the top of the field, would indicate an overestimation of the top stimulus, hence a perspective effect. In order to make the measures for the two positions of the standard comparable to each other, the settings obtained with the standard at the top were translated into scores which represented the size of the top rectangle perceptually equivalent to a 7.5-cm. rectangle at the bottom. This was accomplished by means of the formula $s'/7.5 = 7.5/s$, where s and s' represent, respectively, the match made to the 7.5 standard at the bottom, and the transformed score.

## TABLE 4
### Mean Differences in Size Matches (in Cm.) between Panels Varying in Density and in Randomness

| Groups | H vs. L | NR vs. R |
|---|---|---|
| Grade 1 | .04 | −.18 |
| Grade 4 | .01 | −.17 |
| Grade 8 | .07 | −.17 |
| Adults | .03 | −.18 |
| Combined | .04 | −.18 |

Note.—Negative values indicate greater magnitude of illusion for first-listed panels.

In order to simplify somewhat the analysis of variance (complicated even further, beyond the already rather unwieldy one dealt with for the distance judgments, due to the addition of the variable of the position of standard) each S's settings for the ascending and descending conditions for each stimulus panel were averaged.

The means for each stimulus field at each age level are shown in Table 3, while Table 4 provides a comparison of the H and L panels and the NR and R panels. The analysis of variance of the data falls into two parts. The first, between-Ss portion (comprising the variables of age, order and position of standard, singly and in interaction) discloses no significant source of variance. The second, within-Ss portion (comprising the stimulus variable, singly and in interaction with the others), shows a significant effect due to the stimulus fields ($F = 7.66$, $df = 5/400$, $P < .01$), but no significant interactions.

Summarizing and at the same time elucidating these results, we find the following:

*Stimulus fields.*—While this variable had a significant effect on the size judgments, the results were much less consistent, and less closely in agree-

ment with expectations, than was the case for the distance judgments (cf. Table 3). Duncan's range test shows that the means for both H-R and L-R conditions were significantly higher (at the .01 level) than those for the H-NR, L-NR, and G conditions. This was what had been anticipated, since for these judgments, the lower the score, the larger the apparent size of the top stimulus, and hence the greater the magnitude of the illusion. The results for the C (Control) condition are, however, decidedly out of line, since instead of yielding the highest mean, this condition emerges as intermediate between the two random and the two nonrandom stimulus fields; in fact, the difference between H-R and C is significant at the .01 level (in the wrong direction)! Furthermore, while the two nonrandom-panel means clearly differ from the two random-panel means as expected, the two high-density means are both higher (though not significantly) than the two low-density means, whereas the opposite was anticipated (cf. Table 4).

*Age.*—There appeared to be no consistent differences between the age groups, nor did this variable interact with the stimulus variable.

*Direction.*—Although this variable did not enter into the analysis of variance, inspection of the data shows again a very marked anticipation effect. This effect also tended to decrease with age, but the main difference appeared to be between the adults on the one hand and the three groups of children on the other.

*Position of standard.*—The mean for all judgments made with the variable at the top was 7.44, as against a mean of 7.36 for the transformed settings made with the standard at the top. While this difference was not statistically significant ($F = 2.21$,

$df = 1/80$), it is in the direction of the "error of the standard," involving an overestimation of the standard stimulus per se, which has been encountered previously in the literature (Gardner & Long, 1960; Piaget & Lambercier, 1943). The interaction of this effect with age likewise was nonsignificant; it might be noted nevertheless that the effect appeared to be most marked at the fourth grade, while the eighth graders and adults failed to exhibit any trace of it.

*Order.*—The order-of-judgment variable again failed to affect the judgments significantly.

## Verbalizations Given to the H-R and H-NR Stimulus Fields

Of the 72 school children asked for an interpretation of the H-R and H-NR fields, 43 responded to the H-R and 48 to the H-NR panels. Of these 91 responses, 78 clearly referred to a scene seen in depth, the most common response being a floor, or some variant thereof (e.g., a highway). Interestingly enough, 9 responses made reference to a vertical plane (skyscraper, building), of which 6 came from the eighth graders. Otherwise no notable age differences were found; even the failure-to-respond rate was essentially the same for all groups. Nor were there any very consistent differences between the two panels: the H-NR panel elicited substantially more depth responses on the part of the eighth graders, but the two younger groups gave slightly more depth responses to the H-R panel.

All in all, the results suggest that perspective drawings of this type are effective in conveying a sense of phenomenal depth even to the youngest of the Ss included in this study—a conclusion which is in line with the observed effect of these drawings on the perceptual judgments.

## DISCUSSION

The discussion of the results of this investigation will focus on four separate points:

1. With respect to the distance judgments, the effects of the perspective drawings conformed very neatly to our expectations, showing that information to depth contained in the field, as manipulated in this study, represents a major determinant of the perceptual distortions produced in such drawings. The size judgments, however, yielded much less conclusive results, and did not appear to be consistently related to the stimulus-information variable.

In attempting to account for these somewhat discrepant results, it may be helpful to examine the way in which $S$ would in fact let the stimulus field affect him in making his judgment. It is apparent that for the distance judgments the background forms an essential part of $S$'s field of attention, since his task, to bisect the distance between the two screws, requires him to scan back and forth along a considerable portion of the stimulus field. In the case of the size judgments, on the other hand, it would be much easier for $S$ to ignore the field separating the two stimulus objects to be compared. The incisions made into the field around the objects, and the location of the upper stimulus standing on the top border of the field may also have contributed to the perceptual isolation of the stimuli from the background fields, which would have mitigated their influence. Admittedly, it is difficult to account on this basis for the significant differences between some of the stimulus fields that were found, unless one assumes that, for whatever reason, the intrusion of the field into the size comparisons was especially slight for the two randomly textured panels.

2. The results with respect to the age variable can only be regarded as inconclusive. Certainly there was little suggestion of an increase in the effects of perspective with age, except for the very slight and nonsignificant increase from the first to the fourth grade. It seems, therefore, that if this illusion is a product of learning, whether in the sense of associative or assumptive processes or in the sense of increasing experience in responding to spatial relationships within a field, such learning must run its course fairly early in life. The significant *decrease* in the effects of perspective on the distance judgments for the adults, on the other hand, appear to reflect a more active attempt on the part of these $S$s to counteract this illusion, of which most of them were well aware. In this connection it should be noted that the difference between adults and children may have been at least in part a matter of intellectual level, rather than age, the adults clearly representing a more select group in this respect than the children.

3. An interesting finding was that for both distance and size judgments the control panel itself gave rise to a constant error in the direction of the perspective illusion. A variety of factors could have contributed to this result: the border of the field present in the control panel might by itself have conveyed a sense of depth; perseverative effects from other panels previously exposed might have led $S$ to perceive the control panel in terms of depth; relative height in the field may be interpreted as depth, even in the absence of other cues, just as in other studies (Smith, 1958; Weinstein, 1957) this variable has been shown to represent an effective cue by itself, leading to a certain amount of constancy in size judgments made from photographs.

4. Finally, since one might consider this study as representing a beginning towards an informational approach to the study of space perception, a brief analysis of the possibilities as well as the limitations of such an approach in this area appear to be warranted.

Given the selection procedures employed in the construction of the stimulus fields, the specification of their formal informational content (independent of their role in suggesting depth) is relatively straight forward. As regards the density variable, the two random fields, for which $p$ (the probability of a cell's being filled in) was, respectively, .06 and .20, contain an average of .34 and .72 bits per cell,

respectively. Similarly the two NR panels, for which $p$, for the restricted set of columns from which cells were selected, was .18 and .60, contain, respectively, .69 and .97 bits per cell for each filled column, or .23 and .32 bits per cell for the total field.

The regularity variable, on the other hand, can best be expressed in terms of the redundancy, relative to the corresponding R panels, introduced by the selection procedures. Thus, taking the R panels as a baseline, L-NR would show a redundancy of 32% $(1 - .23/.34)$, while that of H-NR would be 55% $(1 - .32/.72)$. It is interesting to note that for the distance judgments the effect of redundancy was indeed more pronounced for the high-density than for the low-density panels (cf. Table 1).

At first sight, however, the foregoing analysis may appear paradoxical: increasing informational content by increasing the number of cells increases the perspective effect; at the same time, increasing redundancy, which is equivalent to a decrease in informational content relative to the baseline, also increases the effect. In order to resolve this paradox, it is essential to distinguish between the informational content of the stimulus array, which represents essentially the structural complexity of the array, and the "information to depth" provided by such an array. In order to elucidate this point, let us examine more closely the actual role played by the information variable as manipulated here. Increasing the number of cells filled in, or more particularly the proportion of cells, up to .5, increases the effect, by providing the observer with a greater amount of visual information as regards the progressive deformation of the field from the bottom to the top of the panel.[2] Increasing the regularity of the arrangement of the cells by restricting them to a limited set of columns, while decreasing the information, or introducing redundancy, likewise heightens the effect due to the "overdetermination" of the location of the vanishing point, towards which all the columns converge. These considerations underlie our perhaps

somewhat capricious distinction between "amount" and "redundancy" of information, to deal with the density variable and the regularity variable, respectively, as well as our use of the admittedly imprecise term "information to depth."

Speaking more generally, it becomes apparent that any application of an informational model in this area cannot proceed blindly, but must consider the particular ways in which informational content is varied, and their bearing on the perceptual situation and on $S$'s task. Particularly is this true with respect to the role of redundancy, which can probably operate in very different directions, depending on the way it is imposed. For instance, if the restrictions imposed in the selection of cells for the NR fields had involved the rows rather than the columns, the formal amount of redundancy thus introduced would have been the same, yet the effects of this redundancy on the judgments in this task would probably have been less pronounced.

Nevertheless, judiciously applied, the concepts and principles of information theory should prove rewarding in carrying the study of space perception beyond the investigation of isolated cues to the kind of parametric and systematic analysis of the information in the stimulus array on which Gibson has repeatedly insisted. A limited example of this point from our experiment is the treatment of texture density and linear perspective in terms of the more general concept of informational content, which the present approach has made possible.

## SUMMARY

This experiment investigated the effects of different stimulus fields, made up of perspective drawings varying in the amount and regularity of the elements subjected to perspective deformation, on the judgment of relative size and distance in the plane of the drawings. Four age groups, varying from first-grade children to college-age adults, were used as $S$s. The results obtained confirmed the prediction that, as the amount and redundancy of information to depth contained in the field increased, the apparent midpoint

[2] For values of $p > .5$, the figure-ground relations would simply be reversed: the information would in effect be concentrated in the white cells, so that the illusion would decrease again until, for $p = 1.0$, the situation would be formally equivalent to that for $p = 0.0$, i.e., to the control panel. This situation is of course faithfully reflected in the concomitant changes in $H$, the measure of the informational content.

of a segment of a line through the vanishing point would be displaced towards the top of the field. The results for the size judgments were less consistent. The only age difference appeared on the distance judgments, where adults exhibited smaller effects than children between 7 and 14 yr. of age. The implications of the experiment for an informational approach to the study of space perception are briefly considered.

## REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

GARDNER, R. W., & LONG, R. I. Errors of the standard and illusion effects with the inverted-T. *Percept. mot. Skills*, 1960, **10**, 47–54.

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J. A theory of pictorial perception. *Audiovis. commun. Rev.*, 1954, **1**, 3–23.

GIBSON, J. J. Pictures, perspective, and perception. *Daedalus*, 1960, **89**, 216–227.

GLASSER, O. Optical illusions. In O. Glasser (Ed.), *Medical physics.* Chicago: Yearbook, 1944. Pp. 824–827.

PIAGET, J., & LAMBERCIER, M. Recherches sur le développement des perceptions: III. Le problème de la comparaison visuelle en profondeur (constance de la grandeur) et l'erreur systématique de l'étalon. *Arch. Psychol., Genève*, 1943, **29**, 253–308.

SMITH, O. W. Judgments of size and distance in photographs. *Amer. J. Psychol.*, 1958, **71**, 529–538.

SMITH, O. W., SMITH, P. C., & HUBBARD, D. Perceived distance as a function of the method of representing perspective. *Amer. J. Psychol.*, 1958, **71**, 662–674.

WEINSTEIN, S. The perception of depth in the absence of the texture-gradient. *Amer. J. Psychol.*, 1957, **70**, 611–615.

WOHLWILL, J. F. Developmental studies of perception. *Psychol. Bull.*, 1960, **57**, 249–288.

# STIMULUS GENERALIZATION AS A FUNCTION OF UCS INTENSITY IN EYELID CONDITIONING

JOHN J. PORTER[1]

State University of Iowa

One of the implications of the behavior theory developed by Hull and Spence (Hull, 1943, 1952; Spence, 1956), based on studies of classical and instrumental conditioning, is that generalization performance curves for groups at different drive levels will tend to converge. The derivation of this interaction is as follows: $E_s = H \times D_s$, $E_w = H \times D_w$; then, $E_s - E_w = H(D_s - D_w)$, where $D_s =$ strong drive and $D_w =$ weak drive while $E_s$ and $E_w$ are the corresponding excitatory potentials. If $H$ is the habit strength developed to the training stimulus (S) and $\bar{H}$ that developed to the generalized stimulus (S'), then the above derivation implies that a greater difference in excitatory potential will be expected at S than at S'. Thus an interaction is implied.

Studies of stimulus generalization under two levels of motivation in rats by Brown (1942) and in humans by Rosenbaum (1953) yielded converging generalization curves when these authors' time measures were later transformed to speed measures. Newman (1955) studied stimulus generalization in the rat under different drive levels. Although Newman found evidence of convergence with both speed and extinction measures, the interaction was not statistically significant. Both Jenkins, Pascal, and Walker (1958) and Thomas and King (1959) used pigeons to study generalization at different drive levels in the Skinner box. Jenkins et al. (1958) found divergent relative stimulus generalization gradients under conditions of constant drive differences while Thomas and King (1959) found that three of their four drive groups yielded converging stimulus generalization gradients during extinction.

The present experiment measured the predicted performance convergence of high- and low-drive groups with either a 1500-cps or 400-cps tone after all Ss were trained at one drive level with the 1500-cps tone alone. An extinction test was used in order to avoid confounding generalization effects with differential reinforcement effects due to differences in puff intensities.

## METHOD

*Subjects.*—The Ss were 160 women from a course in introductory psychology at the State University of Iowa. Forty-one women were discarded for not meeting a conditioning criterion of more than 8 CRs and less than 36 CRs on Trials 41–80. This conditioning criterion was employed to avoid both ceiling and floor effects when Ss were shifted to either a higher or lower drive condition during the extinction trials. Six women who gave CRs to the CS alone on test trials were discarded as were 7 others who gave 50% or more responses that met the criterion of voluntary responses used in this laboratory (Spence & Ross, 1959). Four additional Ss were discarded due to E error and 2 due to equipment malfunction. The 100 remaining Ss were randomly assigned to four groups.

*Apparatus and recording method.*—The Ss were seated in an adjustable dental chair in a sound shielded room which was illuminated by a shielded 7.5-w. bulb. The Ss' room was separated from the recording and control room by a third intervening room. The equipment used to record eyelid responses was the same as that used in previous studies from the Iowa laboratory (cf. Spence & Taylor, 1951). The CS employed during acquisition was the onset of a 1500-cps tone of 70 db. generated by a Hewlett-Packard audio oscillator and delivered by a 6-in. loudspeaker 4 ft. behind S. During extinction the same 1500-cps tone served as CS for two groups while the remaining two groups received a 400-cps tone of like intensity. The acquisi-

tion UCS was a 50-msec., .6-psi, air puff delivered to the right eye through a .062-in. diameter orifice by a 110-v. ac solenoid valve. During extinction two groups received a .33-psi puff while the other two received a 2.0-psi puff 2500 msec. after CS onset. A dimly illuminated 2.25-in. diameter circular milk-glass disk located 4 ft. in front of S served as a fixation point.

*Procedure.*—Each S was instructed to blink once to the ready signal and then to look at the disk until the tone went off. After the instructions had been read to each S she then received three presentations of the CS alone and one presentation of the UCS alone. The intervals between the ready signal and the onset of the CS were 2, 3, or 4 sec. randomly varied. Intertrial intervals of 15, 20, or 25 sec., given according to a fixed schedule, and averaging 20 sec. were used. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the interval 200–500 msec. following CS onset.

*Experimental design.*—An 80% partial reinforcement schedule was used during training in order to provide for greater resistance to extinction than is obtained with continuous reinforcement. The reinforced and nonreinforced trials followed a prearranged sequence restricted by the provision that no more than 2 nonreinforced trials occurred in each block of 10 trials. On reinforced trials the UCS onset followed the CS by 500 msec.; on nonreinforced trials the UCS began 2500 msec. after the CS onset. McAllister (1953a, 1953b) has shown that little or no conditioning occurs at 2500-msec. intervals, and that the CR extinguishes when the interval is shifted from 500 msec. during acquisition to 2500 msec. during extinction. On both reinforced and nonreinforced trials the CS extended 50 msec. beyond the UCS onset and both CS and UCS terminated simultaneously.

Four groups of 25 Ss were conditioned and extinguished. All groups received identical (.6 psi, 1500 cps) 80% reinforcement conditions for 80 trials followed by 40 extinction trials under one of four conditions. An extinction test was used in order to avoid confounding of the results through further differential performance build-up under the new puff strengths of 2 psi and .33 psi. Group I received 40 extinction trials with a 2-psi puff and the original 1500-cps tone. Group II differed from I only in receiving a .33-psi puff during extinction. Group III received 40 nonreinforced trials with a 2-psi puff and a 400-cps tone. Group IV differed from III only in receiving a .33-psi puff during extinction.

## RESULTS AND DISCUSSION

Performance had stabilized at about 60% responding at the point where extinction began. Inspection of the data and an analysis of variance over the last 10 acquisition trials (71–80) revealed no significant performance differences between the four acquisition groups: therefore these four groups were treated as one group.

Three different response measures were used during extinction and three different blocks of trials were analyzed. Extinction results were analyzed in terms of total number of responses, percentage of responses, and superthreshold excitatory potential ($E_L$) as defined by Spence (1956). The above three measures were taken on Trials 1–10, 1–20, and 1–40, respectively.

Figure 1 presents the percentage of CRs made during Extinction Trials 1–40. The same picture resulted when the other two response measures were used. Examination of Fig. 1 revealed the predicted performance convergence during extinction for the high- and low-drive groups. A similar graphical convergence was obtained on Trials 1–10 and 1–20. Thus there was a smaller difference at the generalized stimulus value than at the original stimulus between the two drive groups. Statistically, such
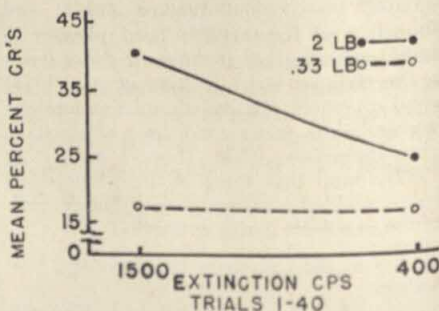


FIG. 1. Mean percentage of conditioned responses for Extinction Trials 1–40.

a condition is represented by a significant interaction between groups. This interaction was significant ($P < .05$, $F = 4.35$ for percentage) over Trials 1–40 for all three performance measures. However this interaction was not significant ($P < .20$) for Trials 1–10 and 1–20 for any of the three performance measures.

Given the interaction over Trials 1–40, the problem becomes one of accounting for the lack of a significant interaction over Trials 1–10 and 1–20 within the framework of the theory used here. As developed earlier, the theory would predict the presence of the sought after interaction within, at the most, a few trials after the change to extinction conditions. Since it required 40 trials to obtain the predicted interaction during extinction it is necessary to examine the effect of inhibition ($I$) on the predicted interaction. In Hull-Spence theory $I$ is assumed to be a function of the number of nonreinforced trials and is assumed to subtract from the quantity ($D \times H$). If $I$ is added to the two equations used earlier to derive the difference equation it may be seen that $I$ cancels out of the difference equation. This leaves $H$ and $D$, or $\bar{H}$ and $D$, as the *differential* factors in this equation. It should be understood that the cancellation of $I$ in the difference equation does not mean that $I$ does not act on performance, but that it does not act differentially. Therefore the interaction predicted by the theory remains unchanged.

Theoretical considerations aside, examination of the range of responses over extinction trial blocks revealed decreasing variability within groups as later extinction trial blocks were examined. This finding suggests that the greater variability of the data during Trials 1–10 and 1–20 may have masked the interaction effect.

## SUMMARY

This study investigated the interaction between drive level and original and generalized stimulus conditions during extinction. One hundred $S$s were first conditioned for 80 trials to respond to a .6-psi air puff on an 80% reinforcement schedule with a CS of 1500 cps. The $S$s were then divided into four groups which were extinguished with either a 1500- or 400-cps tone and a .33- or 2.0-psi air puff. On all nonreinforced trials during acquisition and extinction the UCS was presented, but 2500 msec. after the CS.

The results confirmed ($P < .05$ for Extinction Trials 1–40) the hypothesized interaction between drive level and original and generalized stimulus conditions predicted by Hull-Spence theory. The effect was not significant over Trials 1–10 or 1–20.

## REFERENCES

BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

JENKINS, W. O., PASCAL, G. R., & WALKER, R. W. Deprivation and generalization. *J. exp. Psychol.*, 1958, **56**, 274–227.

McALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 423–428. (a)

McALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422. (b)

NEWMAN, J. R. Stimulus generalization of an instrumental response as a function of drive strength. Unpublished doctoral dissertation, University of Illinois, 1955.

ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

THOMAS, D. R., & KING, R. A. Stimulus generalization as a function of level of motivation. *J. exp. Psychol.*, 1959, **57**, 323–328.

# EXPERIMENTAL EXTINCTION AS A FUNCTION OF NUMBER OF REINFORCEMENTS[1]

## JAMES R. ISON

*University of Rochester*

Overlearning-reversal experiments by Reid (1953), Pubols (1956), and Capaldi and Stevenson (1957) in brightness discrimination and by Pubols (1956) and Ison and Birch (1961) in spatial discrimination have demonstrated facilitation of reversal learning after overlearning compared to control groups reversed at criterion. A similar experiment by Birch, Ison, and Sperling (1960) in differential conditioning demonstrated more rapid extinction of the formerly positive response for the overlearning group and it was concluded that, contrary to the results obtained in the free responding lever press apparatus (Miles, 1956; Perin, 1942; Williams, 1938), resistance to extinction of a running response is not a monotonically increasing function of the number of reinforced trials $(N_g)$. This conclusion was supported, in part, in a runway experiment by North and Stimmel (1960) which demonstrated greater resistance to extinction in a group given 45 reinforcements as compared to groups given 90 or 135. The purpose of the present experiment is to provide further evidence on this relationship in the straight runway apparatus.

## METHOD

*Subjects.*—The *S*s were 75 male hooded rats, approximately 100 days old, obtained from the colony maintained by the Psychology Department of the State University of Iowa.

They were randomly assigned to six groups of 12 or 13 *S*s which received either 10, 20, 40, 60, 80, or 100 rewarded acquisition trials.

*Apparatus.*—A straight alley was housed within a two-unit, black-draped enclosure 4 ft. high, 4 ft. wide, and 11½ ft. long. A plywood panel 4 ft. from the end of the enclosure completely separated the runway section from the goal box except for a hole in the base through which passed the alleyway. The alleyway (covered with glass) was 4 in. high, 3½ in. wide; the start box was 9 in. long, the runway 72 in., and the goal box 18 in. Guillotine retrace doors separated the start box and the goal box from the runway. The entire apparatus was painted flat black and was illuminated by three 75-w. Lumline lights attached to the ceiling of the enclosure directly over the goal box and at distances of 1 and 4 ft. from the start box. These bulbs were screened to give an incident light of approximately 3 ft-c (range, 2.9 to 3.3) on the alleyway. Infrared photobeams permitted the measurement of running speed over two 1-ft. segments of the alleyway beginning 1 and 4 ft. from the start box. The apertures for the lights were ¼ in. in diameter and covered with painted plastic, thus provided little differential stimulation. Mercury switches in the start-box and goalbox doors permitted the measurement of running speed over the entire alleyway. All times were recorded on electronic clocks and the operation of the timing circuits was silent.

*Procedure.*—The *S*s were allowed 10 gm. of food powder each day in wet mash, presented ½ hr. after each daily treatment. They were placed on this schedule and handled for 2 min. each day for 14 days prior to acquisition training. Two trials were given on Days 1 and 2 and 5 trials per day thereafter except for Extinction Day 1, which contained the final rewarded trial followed by 5 nonrewarded trials. The reward was 0.4 gm. of food powder in 0.3 ml. of water, given in a glass dish 4 in. from the goal-box end wall. In extinction the empty glass dish was present, and *S* was detained in the goal box for 30 sec. The minimum intertrial interval in both acquisition and extinction was 18 min. during which *S* was detained in a wooden and

wire mesh box with water available. Extinction was carried to a minimum of 80 trials, continuing if necessary until $S$ took longer than 120 sec. to enter the goal box.

## RESULTS

Four measures are reported, the number of trials to various extinction criteria, running speed in early extinction, the number of avoidance responses made in extinction, and the number of trials to the first avoidance response.

*Trials to criterion.*—A criterion trial was the first trial on which $S$ exceeded a criterion number of seconds to enter the goal box after the start-box door was opened. Four criteria were chosen, 10, 20, 40, and 120 sec., and the mean numbers of trials to each of these are presented in Fig. 1. With the criteria of 10 and 20 sec., the groups did not differ ($F < 1.00$). With the 40-sec. criterion the difference among groups was significant at the .01 level ($F = 3.69$, $df = 5/69$) and on the 120-sec. criterion the difference was significant at the .001 level ($F = 7.48$, $df = 5/69$). On both of these latter criteria the relationship



FIG. 2. Mean running speed in sliding blocks of two trials over the first 3 days of extinction. (The first trial of each day is omitted.)

between $N_g$ and trials to criterion was negative.

*Running speed in extinction.*—In Fig. 2 is depicted the running speed (over the entire alleyway) of the six groups in the first 3 days of extinction. The other speed measures showed essentially identical results. On the initial point Groups 20, 40, 60, and 100 were but little different whereas Groups 10 and 80 were slower. In subsequent trials Group 100 showed the greatest decrement and, with Group 80, was responding at the slowest speeds at the end of the 3 days. Group 10 decreased the least and the other three groups fell roughly in order between Groups 10 and 80. The groups, ranked in mean response speed per trial, were Group 20 ($M = .60$ fps); Group 40, ($M = .55$ fps); Group 10, ($M = .52$ fps); Group 60, ($M = .50$ fps); Group 100, ($M = .41$ fps); Group 80, ($M = .37$ fps). A trial by trial mixed analysis of variance (Lindquist, 1953) yielded a significant Groups effect ($F = 3.64$, $df = 5/69$, $P < .01$), a significant Trials effect ($F = 77.77$, $df = 12/828$,



FIG. 1. The mean number of trials to criteria of 10, 20, 40, and 120 sec. between $E$'s opening the start box and $S$'s entering the goal box.
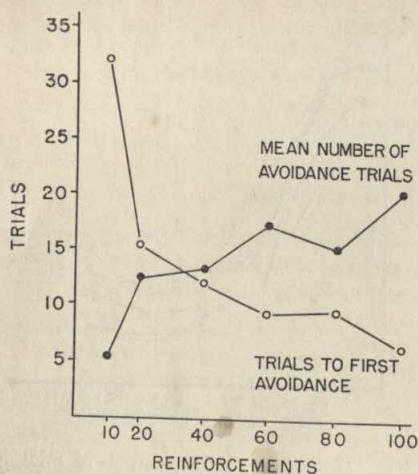
FIG. 3. The mean total number of extinction trials on which an avoidance response occurred and the mean number of trials to the first avoidance response.

$P < .001$), and a significant Groups × Trials interaction ($F = 1.96$, $df = 60/828$, $P < .01$).

*Avoidance responses in extinction.*— An avoidance response was recorded whenever $S$ turned and moved in the direction of the start box. In Fig. 3 two different measures are shown. One is the mean number of trials before the first avoidance response occurred. A simple analysis of variance of these data yielded a significant Groups effect ($F = 4.10$, $df = 5/69$, $P < .005$); the relationship between this variable and $N_g$ was negative. The second is the mean number of trials on which an avoidance response occurred in the 80 extinction trials. An analysis of these data yielded a significant Groups effect, ($F = 5.14$, $df = 5/68^2$, $P < .001$); the relationship between this variable and $N_g$ was positive.

[2] One $S$ was dropped after Extinction Trial 30 because of an experimental error. This reduced the $df$ for this comparison to 68.

## DISCUSSION

Under the conditions of this experiment, trials to extinction criteria of 40 and 120 sec. were negatively related to $N_g$ and running speed in the first 3 days of extinction was nonmonotonically related to $N_g$. These data support the conclusions of Birch, Ison, and Sperling (1960) in their account of overlearning-reversal problems and confirm and extend the findings of North and Stimmel (1960).

This relationship between trials to criterion and $N_g$ is to be contrasted with the negatively accelerated increasing function typically obtained in the Skinner box. One possible reason for this difference is that in the present experiment and in that of North and Stimmel (1960) the reward magnitude ($W_g$) was large; whereas, in the Skinner box the reward was relatively small. Several experiments have suggested that $N_g$ and $W_g$ interact in determining resistance to extinction, the relationship between $R_n$ and $W_g$ being positive at small $N_g$ (Zeaman, 1949) but negative at large $N_g$ (Armus, 1959). This interaction might be reversible, i.e., at large $W_g$, $R_n$ is negatively related to $N_g$; whereas, at small $W_g$, $R_n$ is positively related to $N_g$ over at least part of the range. A second possibility is that the form of the function is determined by some characteristic of the investigated response. The one overlearning-reversal experiment which did not use a running response (McCulloch & Pratt, 1934) found that extended training retarded the subsequent reversal, which is contrary to the results of the later studies. Whether nonmonotone or negative relationships are peculiar to the running response and whether they can be obtained with other responses given appropriate values of $N_g$ and $W_g$ are subject to further investigation.

The positive relationship between the number of avoidance responses and $N_g$ and the negative relationship between trials to the first avoidance response and $N_g$ are consistent with interference theories of extinction stressing the acquisition of frustration-instigated avoid-

ance responses which compete with the approach response (e.g., Birch, 1961; North & Stimmel, 1960). Following Amsel (1958) and Spence (1960), the magnitude of frustration elicited on non-reinforced trials is assumed to be in part a positive function of the number of prior reinforcements, which, with the further assumption that the strength of the avoidance response is positively related to frustration magnitude, is sufficient to account for these two relationships.

## SUMMARY

The relationship between the resistance to extinction of a running response and the number of acquisition trials ($N_g$) was investigated. Six groups of rats received either 10, 20, 40, 60, 80, or 100 rewarded trials followed by 80 nonrewarded extinction trials at five trials per day with an intertrial interval of 18 min. The mean numbers of trials to extinction criteria of 40 and 120 sec. were negatively related to $N_g$, and running speed in early extinction was nonmonotonically related to $N_g$. These data were contrasted with those previously obtained in the Skinner box. In addition, the mean number of trials on which avoidance responses occurred in extinction was positively related to $N_g$ and the mean number of trials to the first avoidance response was negatively related to $N_g$. These latter relationships are consistent with interference theories of extinction which stress the acquisition of competing avoidance responses.

## REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

BIRCH, D. A motivational interpretation of extinction. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1961.* Lincoln: Univer. Nebraska Press, 1961. Pp. 179–196.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

ISON, J. R., & BIRCH, D. T maze reversal following differential endbox placement. *J. exp. Psychol.*, 1961, **62**, 200–202.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCCULLOCH, T. L., & PRATT, J. G. A study of the presolution period in weight discrimination by white rats. *J. comp. Psychol.*, 1934, **18**, 271–290.

MILES, R. C. The relative effectiveness of secondary reinforcers throughout deprivation and habit strength parameters. *J. comp. physiol. Psychol.*, 1956, **49**, 126–130.

NORTH, A. M., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

PUBOLS, B. H. The facilitation of visual and spatial discrimination by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice Hall, 1960.

WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

ZEAMAN, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

# EYE-MOVEMENT LATENCY, DURATION, AND RESPONSE TIME AS A FUNCTION OF ANGULAR DISPLACEMENT [1]

## ALBERT E. BARTZ [2]

### *Applied Research Laboratory, University of Arizona*

Characteristics of eye movements and reaction time to visual stimuli have both been studied systematically since the beginning of the twentieth century. Various components of the visual response were systematically investigated by Dodge and others (Diefendorf & Dodge, 1908; Dodge & Cline, 1901). Using photographic recording techniques, they found that the average latency was about 200 msec., while the eye movement duration was 29 msec. for a 5° movement, and increased to 100 msec. for a 40° movement. Essentially the same results were obtained by Miles (1936) and Hackman (1940).

It must be noted that these time intervals do not reflect the time involved in the process of "seeing" an object in the periphery. After the eye has fixated upon the peripheral stimulus, the observer still must process the new information and make some response.

More recently other investigators have been concerned with the total response time (RT) when there is more than a simple movement in-

volved. Hyman (1953) found that the total RT increased when the task required $S$ to identify the specific location of the stimulus. Words were assigned to various lights and the RT was measured by a voice key set off when $S$ pronounced the correct word for the stimulus location. It should be noted that this increase in RT occurred even though $S$ was not specifically instructed to move his eyes since Hyman's stimulus lights subtended a maximum of only 2.5° of visual angle. This type of RT is more closely related to the problem of seeing, since the total visual reaction must include an identification of what is seen. As was expected Hyman found this identification type of RT to be longer—the lengthening being a function of the statistical probability that a stimulus would appear in the specific location identified. Hyman's vocal RTs varied from 300 to 750 msec.

However, this complex response still does not represent accurately the process of "seeing" an object in the periphery. To see an object in the periphery $S$ must not only identify the location and swing his eyes to it, but also must interpret the stimulus. The present research was designed for two purposes. Experiment I involved the investigation of RTs as functions of angular displacement from the line of regard and the number of stimuli to which $S$ must attend. Experiment II was designed to isolate and measure the various components of the total RT. By using the electrical

method for recording eye movements (Ford & Leonard, 1958; Mowrer, Ruch, & Miller, 1936), it was possible to isolate the latency, eye-movement duration, and the time required for interpreting the stimulus.

## EXPERIMENT I

### Method

*Subjects.*—The Ss were 3 20-yr.-old male volunteer undergraduates. They were free from pertinent visual defects as measured by an Orthorater.

*Apparatus.*—The peripheral stimuli were arranged in a semicircle about S, and were at 40°, 20°, 10°, 5°, and 2.5° right and left. There was also one stimulus at the center or 0°. A point between S's eyes was the center of a circle 6 ft. in radius, and the 11 stimuli were at eye level.

The stimuli were the digits 4, 5, 6, and 9 presented by Burroughs Type BD200S Nixie indicator tubes. The height of each numeral was .305 in., subtending a visual angle of 14' at a distance of 6 ft. These four digits were chosen on the basis of preliminary tests as giving a good response for triggering the voice relay circuit.

The indicator tubes were mounted on a curved panel painted a flat gray to minimize glare. The only illumination in the room was a fluorescent source located 6 ft. behind and 3 ft. above S, giving an illumination on the panel of 2.78 mL.

To keep S looking at the center of the display prior to the presentation of a stimulus, a tracking task requiring continuous monitoring was used. This task required S to follow a light moving in a triangular pattern with another light controlled by a three-button switch. The triangle subtended 1.5° at S's eye, and S could perform satisfactorily while fixated on the center of the triangle.

The S sat in an armchair with a headrest, preventing any horizontal head movements, and wore a headset with an attached carbon microphone. The microphone triggered the voice relay circuit whenever S responded verbally to the number presented in any one of the indicator tubes. This circuit tripped a latching relay that stopped the timer.

The E had controls to select an indicator light at any of the 11 positions and any of the 4 numerals appearing in it.

*Procedure.*—In both the training and experimental sessions, the procedure for the presentation of the stimuli was the same.

The S entered the experimental room which was light-proofed to exclude any extraneous illumination that might reflect from the curved surfaces of the indicator tubes. After the headrest and microphone were adjusted, 10 warm-up trials were given before the session began.

At the start of each trial S began tracking the center display of lights. At intervals of 3, 4.5, or 6 sec. after the start of the tracking task, a number in one of the indicator tubes came on and the tracking lights extinguished. The S then moved his eyes to the position of the stimulus and verbalized the number into the microphone. This response stopped the timer and extinguished the number in the indicator tube. After a 5-sec. rest period the next trial was begun, and S resumed his tracking task. There were two 1-min. rest periods during the experimental session.

*Experimental design.*—To insure reliability of results, all Ss were highly trained prior to the beginning of the experimental sessions. For the training trials the indicator lights in the 20° right and left positions were used. Each S made 144 responses per session, 72 to each position. The four numerals appearing in the indicator lights were randomized among the 72 stimuli for each position. The training sessions were concluded when both the means and SDs became asymptotic. This occurred on Day 16 for 1 S and on Day 19 for the other 2. After the training trials were concluded each S experienced two sessions of responding to all 11 lights.

The experimental trials were run for 12 days and were initiated on the day following the training sessions. In order to determine how RT varies as a function of the number of possible stimuli, it was necessary to divide Exp. I into Sequences A and B. The stimuli in Sequence A consisted of the indicator lights at the 20° and 10° right and left positions. Each of the 16 possible combinations of position and indicator numeral appeared three times in each group of 48 trials, and was randomized throughout each group. The entire session of 144 trials consisted of three of these groups of 48. Sequence A was presented on Experimental Days 1, 2, 11, and 12.

The stimuli in Sequence B consisted of the indicator lights in all 11 positions. To simplify the data reduction the light at the center position was considered as two stimuli, with one-half of the responses counting on the left side of the visual field and the other half counting on the right side. As a result there were 48 possible combinations of position and indicator numeral. Each of the combinations

TABLE 1

Mean RTs and SDs (.01 Sec.) to
Stimulus Positions in
Sequence B: Exp. I

| Position | Direction | | | |
|---|---|---|---|---|
| | Left | | Right | |
| | Mean | SD | Mean | SD |
| 40° | 91.12 | 8.3 | 90.05 | 7.8 |
| 20° | 77.70 | 7.2 | 77.14 | 6.2 |
| 10° | 72.82 | 7.4 | 72.13 | 5.9 |
| 5° | 70.35 | 7.4 | 69.99 | 7.3 |
| 2.5° | 67.79 | 7.8 | 66.81 | 7.7 |
| 0° | 58.43 | 6.9 | — | — |

appeared once in each group of 48 trials and was randomized throughout. Again, each experimental session consisted of three groups of 48 trials. Sequence B was presented on Experimental Days 3 through 10.

*Results*

The data from Sequence B, with all 11 positions, were analyzed by means of a four-factor (Position, Numeral, S, and Day) analysis of variance. All of the main factors were significant beyond the .001 level. Mean RTs to the lights at various positions in the periphery and their SDs are given in Table 1.

The four numerals, or the four vocal responses required (4, 5, 6, and 9), yielded significantly different mean RTs. An examination of the means showed that the vocal responses 4 and 5 were significantly faster than the response 6 or 9. The significant Days effect was due to lengthened RTs occurring on Days 3 and 4.

The Position × Numeral interaction was significant at the .001 level. This indicated that at some positions certain digits yielded faster RTs than at other positions. An inspection of the means showed that in 8 of the 11 positions the fastest mean response was made to the numeral 4. How-

ever, in the other three positions, 40°, 20°, and 10° left, the fastest response was made to the numeral 5.

The significant Position × S interaction indicated that at some positions certain Ss performed better than others. Inspection of the means showed that 1 S was faster at the 40° and 20° positions and slower on the other positions.

The Numeral × S interaction indicated that some Ss responded faster to certain numerals.

The significant Position × Numeral × Day and the Numeral × S × Day interactions indicated that the significant Position × Numeral and Numeral × S interactions varied as a function of the day on which the responses were made. The results of the analysis of variance are shown in Table 2.

The data from Sequence A, with only four positions used (20° and 10° right and left), were analyzed by a similar analysis of variance. The results of this analysis were identical with that of Sequence B, with the exception of the main effect of Day. There was no significant difference be-

TABLE 2

Analysis of Variance of RTs: Exp. I

| Source | df | MS | F |
|---|---|---|---|
| Position (P) | 11 | 10,137.64 | 1,062.65** |
| Numeral (N) | 3 | 1,891.67 | 198.29** |
| S | 2 | 8,683.35 | 910.20** |
| Day (D) | 7 | 113.29 | 11.88** |
| P×N | 33 | 34.21 | 3.59** |
| P×S | 22 | 64.23 | 6.73** |
| P×D | 77 | 10.71 | 1.12 |
| N×S | 6 | 138.83 | 14.55** |
| N×D | 21 | 11.95 | 1.25 |
| S×D | 14 | 61.36 | 6.43** |
| P×N×S | 66 | 15.86 | 1.66* |
| P×N×D | 231 | 7.78 | — |
| P×S×D | 154 | 10.33 | 1.08 |
| N×S×D | 42 | 16.07 | 1.68* |
| P×N×S×D | 462 | 9.54 | |
| Total | 1,151 | | |

* P < .05.
** P < .001.
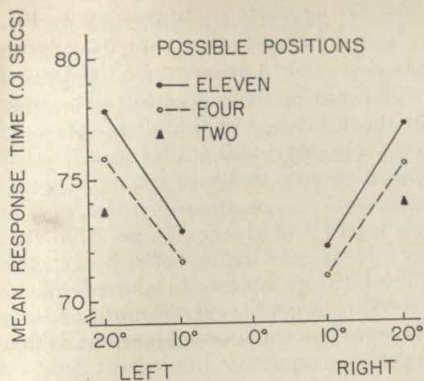
Fig. 1.   Response time as a function of number of possible stimuli.

tween responses made on the 4 days of Sequence A, i.e., Days 1, 2, 11, and 12. This also indicated that the counterbalancing technique used to offset possible transfer effects was successful.

The use of these two sequences was to enable a comparison of responses made to 11 possible positions (Sequence B) as against 4 possible positions (Sequence A). Shown in Fig. 1 is the comparison of the responses at the 20° and 10° left and right positions from both sequences. As is evident from the graph, the mean RTs were faster for the smaller number of possible stimulus positions. All differences were significant at the .05 level. Also shown in Fig. 1 are the data from a situation in which only the 20° left and right positions were used. These means were taken from the last 5 days of the training trials.

## Experiment II

### Method

*Apparatus.*—The same apparatus was used for presenting the stimuli as in Exp. I. To record the eye movements necessary for measuring the components of the total response, electrodes were placed behind external canthi of S's eyes. The output from the electrodes was fed to a Grass Model P-5

preamplifier, and the output of the preamplifier terminated at an oscilloscope. The upper trace of this dual channel oscilloscope was a record of S's eye movements. For the lower trace the input was from the first stage of amplification of the electronic voice key. The sweep was triggered when an indicator light came on. A Dumont oscilloscope camera was used to photograph the tracings.

*Procedure.*—Experiment II was begun on the day following the close of Exp. I. As before, S attended to the tracking task until an indicator light came on, moved his eyes to the stimulus, and responded verbally to the indicator numeral.

*Experimental design.*—Because of the time required to manipulate the camera for recording the CRT trace, it was necessary to reduce the number of responses during the experimental session. It was also necessary to alter the order of presentation of the indicator lights. The order was arranged so that only every third response was recorded by the oscilloscope camera. Since the main interest was in responses to stimuli involving eye movement, only those responses to the 40°, 20°, 10°, and 5° right and left positions were recorded. To insure high reliability in the vocal response, all responses recorded were to the stimulus 5. The remaining three numerals were divided equally among the unrecorded stimuli.

Each S made a total of 96 responses at each session. Of these 32 were recorded, 4 for each of the eight positions. The Ss were not aware that only some responses were being recorded. After 4 days of testing, Ss made
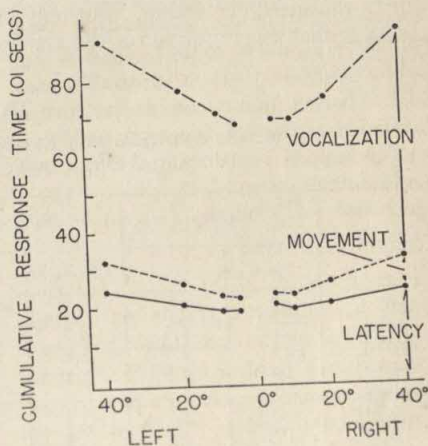


Fig. 2.   Portion of response occupied by three visual components.

a total of 384 recorded responses, or 48 for each of the eight positions.

## Results

The proportions of the total RTs accounted for by the three components of latency, movement, and vocalization are shown in Fig. 2. As was expected, both eye-movement latency and eye-movement duration increased as the angle from the line of regard increased. However, vocalization time (the interval between the cessation of the eye movement and $S$'s vocal response) also increased with angle. The means and $SD$s for latency, duration, and vocalization are shown in Table 3.

No significant practice effect appeared during the 4 experimental days. As in Exp. I there were significant subject differences, appearing mostly in the latency and vocalization.

## DISCUSSION

The type of RT that was investigated in this experiment involved much more than a simple eye-movement latency, so it was logical to expect greater RTs than the latencies reported by Dodge, Miles, or Hackman. Their interest was not in the speed of "seeing," but only in the time that was required for the eyes to begin moving to a peripheral stimulus. In Exp. II it was found that the average eye-movement latency agreed quite well with these previous studies. The over-all mean latency to stimuli at all angles was .213 sec. ($SD = .041$ sec.). This coincides very well with previous data.

However, as mentioned earlier, data on the latency of the ocular reaction does not accurately reflect the process required to see objects in the periphery. The $S$ must get his eyes in motion, swing his eyes to the new object, and then make his response.

It can be seen from Fig. 2 that it is inaccurate to state an "average" value for eye-movement latency, since the time required for the eyes to begin their movement was a function of the angle at which the new stimulus was located. The fact that eye-movement latency increases as the angle from the center line of regard increases was noted earlier by White, Eason, and Bartlett (1962).

It is further apparent that RTs must increase as a function of the angle through which the eyes must move. The actual movement of the eye takes longer as the angle from the center line of regard increases. However, this time interval was extremely small, accounting for only 5% to 10% of the total RT.

A close inspection of the total RT as a function of position shown in Table 1 yielded an interesting observation. The large differences between mean RTs at

TABLE 3

Means and $SD$s for Eye-Movement Latency, Duration, Vocalization, and Total Response in Exp. II

| Position | Latency | | Duration | | Vocalization | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 40° L | 24.8 | 5.3 | 8.7 | 1.0 | 56.0 | 5.9 | 89.5 | 8.5 |
| 20° L | 21.2 | 4.4 | 5.7 | 0.8 | 50.5 | 5.9 | 77.4 | 8.0 |
| 10° L | 19.6 | 2.4 | 3.8 | 0.6 | 47.9 | 5.8 | 71.5 | 9.5 |
| 5° L | 19.7 | 2.5 | 3.1 | 0.5 | 45.6 | 5.7 | 68.5 | 6.5 |
| 5° R | 20.8 | 2.5 | 3.2 | 0.5 | 45.8 | 6.1 | 69.8 | 6.8 |
| 10° R | 19.5 | 2.5 | 3.8 | 0.6 | 46.6 | 6.1 | 69.9 | 6.7 |
| 20° R | 20.7 | 3.9 | 5.7 | 0.9 | 49.0 | 6.7 | 75.3 | 6.9 |
| 40° R | 24.2 | 4.2 | 8.7 | 1.2 | 57.5 | 8.1 | 90.4 | 8.7 |

the various positions could not be fully explained by differences in eye-movement latency or duration. For example, the mean RT to the 40° left position was .911 sec. and that of the 20° left position was .777 sec., for a difference of .134 sec. The difference in eye-movement latency to the two positions was .036 sec., and the difference in eye-movement duration was only .030 sec. Thus latency and movement accounted for less than half of the original difference between the total RTs to the two positions.

Since two of the three components have been accounted for, the difference must be due to the third component, the vocalization time. As defined earlier, vocalization was the time interval between the completion of the eye movement and S's vocal response. As can be seen from Table 3, the time required to make this response after S was looking at the signal varied as a function of indicator light position. In terms of the experimental situation, it took longer for S to "recognize" the numeral presented and verbalize the response when the stimulus was at a greater angle in the periphery.

There are several possible explanations that may be advanced to account for this observation. With the type of recording method used, it was difficult to distinguish very small eye excursions in comparison to the gross movement at the onset of the signal. Therefore, if at the end of a large angular movement, the eye hunts for an exact fixation, this hunting would probably occupy several degrees or less. Small hunting excursions could not be read from the type of records taken. It is possible that this hunting time occupies proportionately more time as the angle increases. This hunting may be comparable to the variable error discussed by Woodworth (1899) in simple motor tasks. He found that variable error for controlled rapid movements increased with amplitude or distance traveled. Fitts (1954) has suggested that in the concept of fixed information-transmission capacity of the motor system, such increased variability is due to decreased information that the

movement provides. It is possible that the hunting may be the variable error of a simple motor response.

Another possible explanation is in terms of accommodation. As the eyes move, the muscles exert different patterns of tension upon the two eyes, and therefore some accommodation may be required in fixating upon a peripheral stimulus. Because of binocular convergence this effect is probably magnified for stimuli close to the eyes. Although the stimuli were 6 ft. from S, sufficient accommodation may be required to make the effect appreciable.

In an experiment mentioned previously (Hyman, 1953), it was reported that RT increased as a function of the probability that a stimulus would appear in a specific location. It would follow from this evidence that RT in general will increase with an increase in the number of possible stimuli to which S must react. As was noted in Fig. 1 the RTs for 11 possible stimuli were significantly longer than for 4 possible stimuli. (A decrease was also noted in the training trials when 2 possible stimuli were used.)

In the 8 days with 11 stimuli, there appeared to be no further learning taking place. With the Days main effect of Sequence A nonsignificant, it can be assumed that no learning in that situation took place over a 12-day period. However, considering the data of Mowbray and Rhoades (1959), it is entirely possible that RTs could be significantly shortened with practice over a long period of time.

Although some investigators (Hackman, 1940) have found shorter RTs to stimuli on the right side, there were no significant differences in this experiment between RTs to stimuli on the left and right sides.

In the general use of the term, the process of "seeing" refers to a wide variety of functions. The usual figure given for RTs to a visual signal is about one-fifth of a second. In such situations the practiced S makes some kind of manual response to the onset of a visual signal. However, a review of the literature shows a wide range of values for

RT—obviously a function of the experimental conditions. In this study, requiring more than a simple reaction, the range in RTs was from .584 sec. for the central stimulus to .906 sec. for extreme positions, with an overall mean of .727 sec.

## SUMMARY

The present research was initiated with two purposes in mind: (a) to determine the speed of seeing in a complex visual task (Exp. I), and (b) to isolate and measure the various components of the total response (initial latency, travel time of the eye, and the response time for interpreting the signal).

Results of Exp. I showed that RT increased as the angle from the center line of regard increased. There was no significant difference between pairs of means for right and left sides. It was also found that response time increased as the number of possible signals increased. In Exp. II, the time required for each of the three components of the response increased as the angle increased. Several interpretations of the positive relationship between angle and the time required for S to make his vocal response after his eyes had reached the signal were considered.

## REFERENCES

DIEFENDORF, A. R., & DODGE, R. An experimental study of the ocular reactions of the insane from photographic records. Brain, 1908, 31, 451–489.

DODGE, R., & CLINE, T. S. The angle velocity of eye movements. Psychol. Rev., 1901, 8, 145–157.

FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. J. exp. Psychol., 1954, 47, 381–391.

FORD, A., & LEONARD, J. L. Techniques for recording surface bioelectric direct currents. USN Electron. Lab. res. Rep., 1958, No. 839.

HACKMAN, R. B. An experimental study of variability in ocular latency. J. exp. Psychol., 1940, 27, 546–558.

HYMAN, R. Stimulus information as a determinant of reaction time. J. exp. Psychol., 1953, 45, 188–196.

MILES, W. R. The reaction time of the eye. Psychol. Monogr., 1936, 47 (2, Whole No. 212).

MOWBRAY, G. H., & RHOADES, M. V. On the reduction of choice reaction times with practice. Quart. J. exp. Psychol., 1959, 11, 16–23.

MOWRER, O. H., RUCH, T. C., & MILLER, N. E. The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements. Amer. J. Psychol., 1936, 114, 423–428.

WHITE, C. T., EASON, R. G., & BARTLETT, N. R. Latency and duration of eye movements in the horizontal plane. J. Opt. Soc. Amer., 1962, 52, 210–213.

WOODWORTH, R. S. The accuracy of voluntary movement. Psychol. Monogr., 1899, 3 (2, Whole No. 13).

# SUPPLEMENTARY REPORT: DIRECTION OF CHANGE IN CS IN EYELID CONDITIONING [1]

FRANK A. LOGAN AND ALLAN R. WAGNER

Yale University

The assumption that the important parameter of the CS is the amount of change from the pre-CS condition to the CS condition (e.g., Logan, 1954; Perkins, 1953) implies that a decrease in intensity should be as effective a CS as the corresponding increase in intensity. The assumption that the absolute value of the CS has a motivational (dynamogenic) property (e.g., Hull, 1952) implies that an increase in intensity should be more effective. Kish (1955) found tone-off to be a less effective CS than tone-on for avoidance conditioning in rats but Schwartz and Goodson (1958), using a comparable situation, found these events to be equally effective. Hansche and Grant (1960) concluded that light-off was as effective as light-on for eyelid conditioning under a procedure in which the light was off between trials for all Ss. The present study compares an increase with a decrease in intensity between two nonzero values treated symmetrically.

*Method.*—The general features of the eyelid conditioning apparatus, recording equipment, and procedures have been described elsewhere (Dufort & Kimble, 1958). The CS was provided by a circular milk glass disk, 2.25 in. in diameter, set in a flat black ground, and illuminated from behind by General Electric NE30 neon bulbs. The onset of the CS was either an increase from two to four bulbs or a decrease from four to two bulbs. In each case, the CS intensity lasted for 600 msec. during the last 100 msec. of which a 2-lb. air-puff was delivered to the corner of the eye. The non-CS intensity remained on during the intertrial interval which averaged 20 sec. in length.

Five test trials of CS or UCS alone were followed by 60 conditioning trials. During these trials the CS for half of the Ss was an increase while for the other half it was a decrease in illumination. All Ss were then given 20 additional conditioning trials with the opposite CS. The results from 16 female student nurses were combined with those from 40 male undergraduates since they were virtually identical.

*Results and discussion.*—The results are shown in Fig. 1. Both the increase and de-
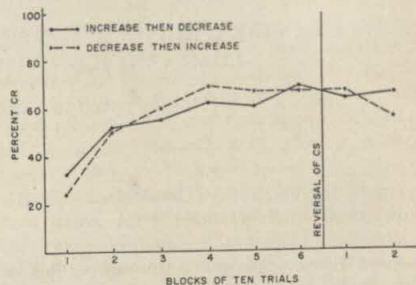


FIG. 1. Percentages of conditioned responses during training. (The solid curve refers to Ss for whom the CS was an increase from the between-trials intensity during the first 60 trials and a decrease during the next 20 trials, while the dashed curve refers to Ss who received these CS conditions in the reverse order.)

crease in intensity were clearly and equally effective CSs in producing a relatively high level of conditioning. Although the null hypothesis cannot be proven statistically, the standard error of the difference between the groups at the end of training was only 7% and hence it is unlikely that the true difference deviates very much from zero. The data thus indicate the greater relative importance of the change parameter of the CS rather than its absolute intensity.

The degree of transfer when the direction of change was reversed is remarkable. Indeed, a slight drop in performance would be expected because of the "extinction trial" given inadvertently when the non-CS intensity was reversed between the last acquisition trial and the first reversal trial. This finding suggests that generalization should be viewed in terms of a surface including the non-CS condition as well as the CS condition. However, it will require a large parametric study adequately to characterize this surface.

## REFERENCES

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, 56, 1–7.
HANSCHE, W. J., & GRANT, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, 59, 19–26.
HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
KISH, G. B. Avoidance learning to the onset and cessation of conditioned stimulus energy. *J. exp. Psychol.*, 1955, 50, 31–38.

LOGAN, F. A.  A note on stimulus intensity dynamism
(V).  *Psychol. Rev.*, 1954, 61, 77–80.
PERKINS, C. C., JR.  The relation between conditioned
stimulus intensity and response strength.  *J. exp.
Psychol.*, 1953, 46, 225–231.

SCHWARTZ, M., & GOODSON, J. E.   Direction and rate
of conditioned stimulus change in avoidance per-
formance.  *Psychol. Rep.*, 1958, 4, 499–502.

# SUPPLEMENTARY REPORT: THE EFFECT OF STIMULUS DURATION AND LUMINANCE ON VISUAL REACTION TIME [1]

DAVID RAAB AND ELIZABETH FEHRER

*Brooklyn College*

Raab, Fehrer, and Hershenson (1961) found that simple reaction time (RT) was independent of stimulus duration over the range of 10 to 500 msec. Luminance, on the other hand, was found to be an important determiner of RT. Since intensity rather than total energy (intensity times duration) determined RT, it is obvious that the critical duration (CD) for RT is 10 msec. or less for the three luminance levels (3000, 30, and 0.3 ft-L) investigated.

The term critical duration has been borrowed from visual threshold studies, which have shown reciprocity (Bunsen-Roscoe law) up to a CD of approximately 100 msec., beyond which temporal integration ceases and the threshold is defined solely in terms of luminance.

It seemed worthwhile to determine the CDs in the mediation of RT for the luminances previously studied and the relation between RT and stimulus duration below these critical values. In the experiment to be reported, the six durations ranged from 0.5 to 20 msec., and thus overlapped the range used previously. Two additional intermediate luminances were included.

*Method.*—Target flashes were generated and RTs measured by the same equipment as that employed in our previous study. A single Tektronix wave-form generator provided the gating pulses for the glow modulator tube; pulse durations were switched between trials, as required.

In order to generate flashes having wave forms as rectangular as possible, the driving pulses were shaped to "overvolt" the glow modulator tube, and the tube itself was placed next to an ultraviolet source. With these arrangements, flash energy was found to be

proportional to flash duration within 0.5 db. from 0.5 to 20 msec. The circular target, 1 cm. in diameter, subtended 1° 10' of arc and was viewed binocularly.

Two senior honors students and the 2 authors served as *S*s. Each *S* served in 30 experimental sessions. Computations are based on data of the last 25 sessions. Only one luminance was used in a given session; the five luminances were counterbalanced over test days for each *S*. Each session consisted of four blocks of 18 trials each, in which each combination of the six durations and three foreperiods appeared once in random order. Only the four longer durations could be explored for the 0.3 ft-L luminance, since this light was below foveal threshold when presented for 0.5 or 1 msec.

Each session began with 5 min. of dark adaptation. Four practice trials preceded the recorded trials. The four blocks were separated by 1-min. rest periods.

The 12 RTs obtained in a session for each of the six durations were reduced to 10 by
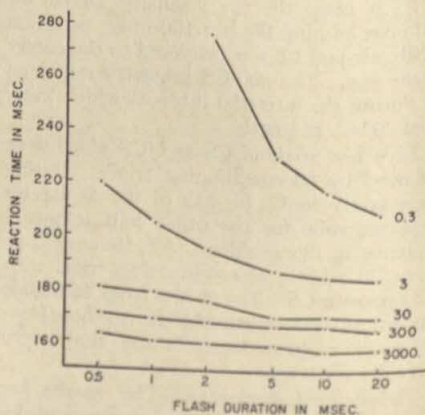


FIG. 1.  Reaction time as a function of stimulus duration. (The parameter is flash luminance in ft-L. Each data point is the mean for 4 *S*s.)

discarding the longest and the shortest RT. Testing over 25 days (5 at each luminance) thus yielded means for each luminance-duration combination based on 50 trials.

*Results and discussion.*—Mean RTs for the 4 *S*s combined are plotted in Fig. 1. Each data point is thus based on 200 RTs

For the two highest luminances, duration is unrelated to RT over the range studied. For the 30 ft-L flash, there was a 10-msec. increase in RT when its duration was reduced from 5 to 0.5 msec. For the two lowest luminances, stimulus duration has a far more marked effect on RT, RT being obviously an accelerated function of flash briefness.

Our results show that the CD for moderately intense stimuli (3000 and 300 ft-L) is remarkably brief, being less than 0.5 msec. At 30 and at 3 ft-L, CD lies between 2 and 5 msec. For the weakest target, the CD lies between 10 and 25 msec. The present study shows a small decrease in RT as duration increased from 10 to 20 msec. In the previous study, a smaller decrease occurred between 10 and 25 msec., but there was no further decrease when this stimulus was prolonged beyond 25 msec.

These CDs for RT are far shorter than the 100-msec. value previously reported for absolute threshold (e.g., Baumgardt & Hillmann, 1961) or the minimal value of 30 msec. reported by Graham and Kemp (1938) for the incremental threshold at their highest background luminance. The three dependent variables, RT, RL, and DL, are thus differently related to stimulus duration, with the CD being obviously shortest for RT.

Although luminance differences are confounded with test days (i.e., only one luminance was studied in a given test session), the effect of luminance on RT is pronounced and is apparent at all durations studied. That RT decreases when luminance is increased is consistent with earlier findings (see Woodworth & Schlosberg, 1954). But the form of the relation between luminance and RT will depend on stimulus duration *unless* each stimulus duration is greater than the CD. In other words, our data could be replotted to display six different luminance-RT functions, one for each flash duration.

Our results show that although the overt response to a target flash may not appear until much later, the minimal latency of that response is determined very shortly after stimulus onset. The finding that increasing duration may cease to be effective long before the criterion response appears parallels the classical observation of this fact made by Hartline (1934). The fact that RT is determined by so brief a "package" of luminous energy is consistent with our earlier finding that RT is independent of the growth (with duration) of phenomenal brightness. In addition, it helps to explain why retroactive (metacontrast) masking of a flash does not affect its RT (Fehrer & Raab, 1962).

REFERENCES

BAUMGARDT, E., & HILLMANN, B. Duration and size as determinants of peripheral retinal response. *J. Opt. Soc. Amer.*, 1961, 51, 340–344.
FEHRER, E., & RAAB, D. Reaction time to stimuli masked by metacontrast. *J. exp. Psychol.*, 1962, 63, 143–147.
GRAHAM C. H., & KEMP, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, 21, 635–650.
HARTLINE, H. K. Intensity and duration in the excitation of single photoreceptor units. *J. cell. comp. Physiol.*, 1934, 5, 229–247.
RAAB, D., FEHRER, E., & HERSHENSON, M. Visual reaction time and the Broca-Sulzer phenomenon. *J. exp. Psychol.*, 1961, 61, 193–199.
WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

# SUPPLEMENTARY REPORT: MEANINGFULNESS AS A DIFFERENTIATION VARIABLE IN THE VON RESTORFF EFFECT [1]

HAROLD ROSEN, DONALD H. RICHARDSON, AND ELI SALTZ

*Center for the Study of Cognitive Processes, Wayne State University*

The present study was designed to test an aspect of the theory for the von Restorff effect proposed by Saltz (1960). This theory defines a *differentiation* construct in terms of two variables: (*a*) Similarity between an item

and other items in a list, and (*b*) amount of prior reinforcement (e.g., "familiarization") of the item. Isolation techniques typically involve differentiation of an item by reducing its similarity to other items in the list. However, if all the items in a list are already highly differentiated, reduction of similarity would be expected to have a smaller effect than if

TABLE 1

Mean Intralist Rank and Mean Number of
Correct Anticipations of the Isolated
and Control Terms over 15
Learning Trials

| List | Item | Ranks | | Correct Anticipations | |
|------|------|-------|------|-------|------|
| | | Mean | SD | Mean | SD |
| Low *m* | Isolated | 4.53 | 2.24 | 5.73 | 3.53 |
| | Control | 6.53 | 1.62 | 2.12 | 2.95 |
| High *m* | Isolated | 5.00 | 1.63 | 7.70 | 3.94 |
| | Control | 5.45 | 1.89 | 6.48 | 3.75 |

the other items in the list were low in differentiation.

This deduction was tested by comparing the effects of isolation of an item in a serial list of high-meaningfulness items with the effects obtained by isolating an item in a serial list of low-meaningful items, since Noble (1953) has demonstrated that meaningfulness is related to amount of prior experience with an item. Isolation was accomplished by typing one word of the list in red, all the other words being typed in black.

*Method.*—The *S*s were 132 students in introductory psychology at Wayne State University, and were randomly assigned to conditions except that *N*s of conditions were made equal.

Two basic lists were used, each consisting of nine items taken from Noble's (1952) meaningfulness scale. One list consisted of low-meaningful items (*m* range = 1.05 to 1.50): MEARDON, BYSUSS, VOLVAP, LATUK, GOKEM, POLEF, SAGROLE, WELKIN, NARES. The high-meaningful items (*m* range = 7.39 to 9.61) in a second list were: INSECT, JEWEL, HEAVEN, OFFICE, WAGON, DINNER, MONEY, ARMY, KITCHEN. The item in Position 5 was typed in red and served as the isolate in the experimental condition. All other items were typed in black. The control lists were typed entirely in black. Each of the two basic lists was organized into six different random orders, with a different term serving as isolate in each order. An *S* learned one of these orders as a serial list.

Items were exposed on a Lafayette memory drum at a 2-sec. rate with a 4-sec. intertrial interval. After an initial trial in which *S* pronounced aloud the items in order of their appearance, *S* was given 15 anticipation trials.

*Results.*—The relative effects of isolation for high- and low-meaningful material were evaluated by determining the effect of isolation on each *S*'s serial position curve. For each *S*, the items in the list were ranked from greatest number of correct anticipations in

15 trials (Rank 1) to least correct anticipations (Rank 8). The mean ranks of the isolated and control items, summarized in Table 1, indicate that isolation has a much greater relative effect in a list of low-*m* than a list of high-*m* items. The main effects due to isolation are significant beyond the .001 level ($F = 14.36$, $df = 1/128$) and the Isolation × Meaningfulness interaction is significant beyond the .025 level ($F = 5.77$, $df = 1/128$). Table 1 shows that the interaction effect also occurs for mean number of correct anticipations over 15 trials. However, the anticipation data are less persuasive than the rank data, since the anticipation data could be an artifact due to the rapid learning of high-*m* items, both isolated and control. In brief the original prediction based on the differentiation formulation is sustained.

Saltz and Newman (1959) found that isolation resulted in an increased tendency for the isolated term to be emitted as an intrusion response to other stimuli in the list. The intrusion data in the present study were analyzed by means of a chi square test since the data were markedly skewed: the median number of intrusions was zero in the low-*m* control groups, and close to zero in the other groups. For low-*m* lists, 58% of the *S*s in the isolated condition emitted the isolated term as an intrusion at least once during learning; only 33% of the control *S*s emitted the control item. This difference is significant ($P < .05$). The corresponding percentages in the high-*m* conditions are 58% and 67%, producing a nonsignificant difference opposite in direction from that for low *m*. Saltz (1960) has hypothesized that differentiation increases the tendency for a response to be emitted, and that this tendency is basic to the relatively high intrusion rate for the isolated term. In terms of this position, the fact that the high-*m* isolation condition produces no greater tendency toward intrusions than the high-*m* control condition is consistent with the position stated previously in this paper: in a list of highly meaningful items, the terms are already relatively differentiated, and so isolation will contribute relatively little additional differentiation.

REFERENCES

Noble, C. E. An analysis of meaning. *Psychol. Rev.,* 1952, 59, 421–430.
Noble, C. E. The meaning-familiarity relationship. *Psychol. Rev.,* 1953, 60, 89–98.
Saltz, E. Similarity and differentiation in verbal learning. Paper read at Psychonomics meetings, Chicago, 1960.
Saltz, E., & Newman, S. E. The von Restorff isolation effect: Test of the intralist association assumption. *J. exp. Psychol.,* 1959, 58, 445–451.

# Journal of

# Experimental Psychology

## USE OF TEMPERATURE STRESS WITH COOL AIR REINFORCEMENT FOR HUMAN OPERANT CONDITIONING [1]

GORDON L. PAUL, CHARLES W. ERIKSEN, AND LLOYD G. HUMPHREYS

*University of Illinois*

The attempts that have been made to investigate operant conditioning in normal human Ss have dealt primarily with verbal behavior. The general approach has been to manipulate the frequency of usage of word classes using as a reinforcement various signs of social approval such as "umhm," "good," etc. However, Buchwald (1960) has raised doubts as to the effectiveness and unambiguity of these signs of social approval as reinforcers and other Es (Eriksen, 1960; Krieckhaus & Eriksen, 1960; Levin, 1961) have found that changes in S's verbal behavior do not convincingly occur in the absence of S's ability to verbalize relevant intentions and hypotheses concerning relationships between the experimental variables and changes in his behavior.

Attempts to obtain operant conditioning of nonverbal responses have been reported by Verplanck (1955, 1956) and Hefferline, Keenan, and Harford (1959). However, Verplanck's results have come under seri-

ous question as a result of an attempt to replicate his procedures reported by Azrin, Holz, Ulrich, and Goldiamond (1961) and the results of Hefferline et al. are difficult to interpret owing to inadequate questioning of Ss for relevant verbalizable hypotheses.

Interpretation of these studies is also complicated by the failure of the investigators to include appropriate control groups. To show clearly that operant conditioning occurred, these studies would have required a control group, treated identically with the experimentals and receiving the same number of reinforcements with the exception that the reinforcements for the controls would have been randomly administered and not contingent upon any particular response.

In the study reported below we have attempted to remedy the deficiencies in the previous attempts at operant conditioning of nonverbal responses in human Ss. By the use of a new technique we were able to administer positive primary reinforcement to human Ss under conditions directly comparable to those obtaining in animal experimentation. Three

329

different classes of motor responses were studied and particular care was taken to assess the extent to which behavior modification depended upon awareness as signified by $S$'s ability to verbalize relevant mediational steps that intervened between the experimental variables and his behavior.

## METHOD

*Subjects.*—Fifty-two females ranging in age from 18 to 30 yr. were obtained on a volunteer basis from elementary classes in psychology, education, and speech during the summer session at the University of Illinois.[2] As an added inducement, 5 dollars in services at a local beauty salon was offered volunteers.

*Procedure.*—Subjects were run individually. After $S$ had changed from street clothes to a playsuit or swimsuit she was told that this was an experiment on the effects of heat and isolation on psychological processes. Further, that the study was government supported and was being conducted as part of a "lady in space project." The $Ss$ were then shown a commercial speed temperature graph demonstrating the extremely high ram air temperature (1500°) recorded on the surface of rockets. The graph also stated that "experiments by the University of California at Los Angeles have shown that man can withstand a temperature of 250°." They were then told that since this study was only concerned with psychological reactions, an increase in relative humidity had been substituted for an increase in temperature above a certain point in order to remove any physical risk. During this preliminary period $E$ also determined whether $S$ had any physical restriction which would be a basis for exclusion from the study and then gave $S$ a brief synopsis of what she would do for the rest of the period.

Next, $S$ was instructed in the operation of a stationary ergocycle which she rode for 5 min. She was told the purpose of this exercise was to bring her body temperature up so that it would not be such a drastic change when she entered the "space" chamber. After riding the ergocycle $S$ was immediately directed into the chamber (maintained at a temperature of 105° F. and relative humidity

of 85%). She was told to sit with her face directly in front of an insulated pipe which protruded into the chamber since cool air, high in oxygen content, would be blown through the pipe from time to time.

The chamber door was then closed and bolted. Further conversation was by means of an intercommunication system with $S$ being observed through a one-way mirror in one porthole of the chamber. The $S$ was told that she should ask any questions she might have immediately following the instructions, since no communication could take place once the "sonar" signals had begun. A standard set of instructions was then read to $S$ further explaining the "increasing" heat in the chamber. The $S$ was told: "You will probably find the air most beneficial if you breathe it directly." Also included were detailed instructions on the performance of a pseudo task consisting of differentiating between sonar signals and forming patterns with colored washers on a peg board: "Make as many different designs as you can. If you drop a washer, replace it with one of appropriate color from the stack on your right. Completely fill one-half of the board and then work back to the other half. Continue this process until I tell you the time is up. Are there any questions?"

Approximately 5 min. elapsed from the time $S$ first entered the chamber through completion of the instructions. Following the instructions a time clock and tape recorder which presented the signals for the pseudo task were started. The $E$ observed $S$ for the following 5 min. and determined the specific response within the various response classes which was to be conditioned.

The recording apparatus was then started and the initial rate of the operant was recorded by throwing a toggle switch each time the selected response occurred. After 5 min. of recording the operant level, a switch was thrown which connected a programmer and timer into the circuit marking the beginning of the conditioning phase. During this phase which lasted for 35 min., 10 sec. of cool air (primary reinforcement) and a red light (secondary reinforcement) were automatically timed and administered on a programed schedule activated by the same switch which recorded responses.

A decreasing ratio reinforcement schedule was used in which the first 20 responses received 1:1 reinforcement, the next 15 responses received 2:3 reinforcement, the next 14 received 1:2 reinforcement, and the remaining responses received 1:3. The latter schedule applies to responses in which only

---

[2] Female $Ss$ were used since several males (run on a pilot basis), to demonstrate the cultural stereotype of "masculinity," refused air although they greatly desired it. This confounding was rare in females.

one program step could be made in every 10-sec. period. The record of responses on the other hand was kept on each specific response regardless of time interval with the exception of single responses lasting for a period of more than 10 sec. which were then recorded as two responses. This procedure was adapted to maximize unawareness.

Following the 35 min. of conditioning, the programmer and timer were disengaged and a 10-min. extinction phase was initiated in which responses were recorded as above.

Upon completion of the conditioning phase, S was assisted out of the chamber and escorted directly to a shower room, with instructions to return to the laboratory for a short interview.

Control Ss were tested for each response class, receiving exactly the same treatment as experimental Ss with the exception of the conditioning phase. During this 35-min. period a technician operated a separate switch which activated the programmer and timer that delivered rewards. These rewards were administered in accordance with a master program which had been prepared from the average number of program reinforcements over time that had been delivered to the experimental Ss in the same response class. Thus the control Ss received essentially the same number of rewards at the same times as the experimental Ss but the rewards were not contingent upon any specific response of theirs. The record of responses was taken in the same manner for both experimental and control Ss.

Upon S's return from the shower room the following questions were asked and answers recorded verbatim:

1. Would you tell me, in your own words, exactly what we're studying?

2. Did you like or look forward to the cool air and oxygen? Relatively how much?

3. Were you able to concentrate on the task the entire period?

4. How well do you believe you did?

5. Do you feel you had control of the cool air at any time? How? When?

6. Would you mind telling me specifically what you thought about when you were in the chamber.

7. Did anything you did at any time during the period have any influence upon when the light or air came on? When?

8. If I were to tell you that something you did determined when the air came on what would you guess it might be?

9. Would you describe any emotions you may have experienced during the period.

Running notes of S's behavior in the chamber were kept, specifically noting actions correlated with the operant or onset of reward. A record was also kept of S's performance on the pseudo task to provide some "feedback" and in combination with the buffer items in the postexperimental interview to maintain face validity of the "space" experiment.

The reliability of E in recording responses was determined for each of the three response classes by having 2 other Os, the environmental laboratory technician and a graduate student in psychology, simultaneously observe Ss for an entire period and record the operant responses with paper and pencil by 1-min. intervals.

Three general categories of responses were studied: hand movements (Group H), face and mouth movements (Group FM), and foot movements (Group F). Subjects were randomly assigned to one of the response class groups with the specific response in this category that was to be reinforced (e.g., Group H, touch chin with right hand; Group FM, press lips; Group F, tap ball of right foot on deck) determined during the 5-min. observation period that preceded the determination of the operant level of the response. Twelve Ss were run in each response category with an additional 4 control Ss for each category. In the course of experimentation an additional 4 Ss were excluded: 2 because the experimental period had to be terminated before completion; 1 because she did not like the reinforcement; and 1 who immediately became aware of the contingency of the reinforcement and her response and stopped responding because she "thought it would ruin the experiment." No Ss were initially aware of the true nature of the study.

*Apparatus.*—The experiment was conducted in the Physical Environment Laboratory at the University of Illinois. The laboratory housing all equipment was maintained at a temperature of 70° F. and 35% relative humidity.[3]

The ergocycle which Ss rode prior to entrance into the chamber was the Illinois electrodynamic bicycle ergometer with an armature load of 2400 ft-lb at 50 rpm. The chamber was a 6-man low pressure chamber soundproof and cork insulated with cylindrical

[3] Lawrence Siler, technician in charge of the Environmental Unit, maintained a vigilant watch on all equipment used, assuring constant environmental conditions throughout the experiment.

TABLE 1

RELIABILITY COEFFICIENTS BETWEEN $E$
AND 2 $O$s BASED ON SUCCESSIVE
5-MIN. PERIODS

| Group | $EO_1$ | $EO_2$ | $O_1O_2$ |
|-------|--------|--------|----------|
| H | .999 | .998 | .999 |
| FM | .947 | .915 | .947 |
| F | .972 | .923 | .965 |

external dimensions of 12 × 7 ft., built for
the United States Air Force by the Pitts-
burgh Des Moines Steel Company. The
internal compartment was a 6-ft. sphere.
The environmental controls were of the
Johnson Service Company. Conditions inside
the chamber were maintained at a tem-
perature of 105° F. (+ or −2°) and relative
humidity of 85% (+ or −5%). The air
velocity within the chamber determined by
an aneotherm air meter varied from 10 to 18
ft. per min. at different points.

Rewards were administered through a 1-in.
insulated pipe, 4 ft. long. Twin blowers drew
in air from the laboratory (70° F. and 35%
humidity) and blew it through the pipe,
emerging at $S$'s face at 78° F. with a velocity
of 710 ft. per min. Operant responses were
recorded on a Hunter stylus recorder which
was wired directly to the control switch.
The mechanism which timed and administered
rewards, also wired in series to the control
switch, included a Hunter Model III elec-
tronic timer and a Ridgely automatic pro-
grammer. A 25-w. red light inside the cham-
ber was wired to operate simultaneously with
the activation of the blowers and to serve as a
secondary reinforcer filling the short time
interval between activation of the blowers
and arrival of the cool air on $S$ (less than 1
sec.).

The apparatus for the pseudo task con-
sisted of a 1 × 3 ft. pressed wood board half
of which was painted white with 42 bolts
mounted on its face. The bolts contained
seven each of three different colored washers.
The sonar signals consisted of a magnetic
tape of variable signals ranging between 800
and 2000 kc. as recorded from a variable
oscillator. The tape was presented over a
tape recorder which was wired into the
intercommunication system with which the
chamber was equipped.

## RESULTS

Before turning to the results the
reliability of $E$ in recording the oc-

currence of responses needs to be
documented. Table 1 presents the
product-moment coefficients obtained
between 2 independent $O$s and $E$.
These figures were based on a single $S$
from each experimental group with
the total responses in successive 5-
min. periods as the basic unit.

On the basis of $S$s' verbalizations
during the postexperimental inter-
view, they were classified into aware
and unaware subgroups within each
response category immediately upon
completion of questioning. Those $S$s
were classified as aware who gave an
affirmative answer to Questions 5, 7,
or 8 in the postexperimental interview
and who could further state the
correct contingency between the oper-
ant response and the reward, or name
a response the occurrence of which
would have led to an increase in the
operant. In the case of these latter
correlated hypotheses the accuracy of
$S$'s verbalization was checked against
the running notes that had been made
of her behavior during the experiment.
Subjects who gave negative answers
to these three questions or who could
not state the correct or correlated
contingency were classified as un-
aware. Table 2 shows the number of
aware $S$s in each of the three response
groups and also the question of the
three listed which elicited the ver-
balization of the correct or correlated
contingency.

TABLE 2

NUMBER OF AWARE $S$s IN EACH GROUP
DESIGNATED BY THE QUESTION
ELICITING EVIDENCE OF
AWARENESS

| Group | Question | | |
|-------|---|---|---|
| | 5 | 7 | 8 |
| H | 3 | 2 | 1 |
| FM | 1 | 2 | 4 |
| F | 0 | 1 | 2 |

Evidence of conditioning was evaluated separately within each of the three response groups by a simple analysis of covariance (aware, unaware, and control subgroups) using the average number of responses during the last 15 min. of the conditioning phase for each subgroup adjusted according to the average number of responses occurring during the operant period. The summary of these covariance analyses is given in Table 3. The adjusted and unadjusted means are presented in Table 4.

These analyses indicate that conditioning was obtained in Groups H and FM but, as seen in Table 4, apparently only in Ss who became aware of the contingency between their response and the reward.

Curves of the average response rate per minute over successive 5-min. periods are shown in Fig. 1 for aware, unaware, and control Ss within each of the response groups. These plots reveal, without adjustment for operant level, that the two groups demonstrating conditioning show a steady increase of response for the aware Ss during conditioning and a decline

during extinction. No such change is present for the unaware Ss or the controls.

There was some concern that variations in the daily environment of Ss might have some effect on the results since the value of the reward depended for the most part upon the effect of heat and humidity within the chamber. Temperature and humidity data for the hour on which each S arrived for the experiment were obtained from the Illinois State Climatologist. The ranges of the mean temperature and

### TABLE 3

COVARIANCE ANALYSES ON THE MEAN RESPONSES DURING THE LAST 15 MIN. OF CONDITIONING ADJUSTED TO OPERANT LEVEL

| Group | Source | df | MS | F |
|-------|--------|-----|--------|--------|
| H | Subgroups | 2 | 446.9 | 4.06* |
| | Error | 12 | 110.1 | |
| FM | Subgroups | 2 | 1463.7 | 7.93** |
| | Error | 12 | 184.5 | |
| F | Subgroups | 2 | 43.7 | .18 |
| | Error | 12 | 237.5 | |

\* $P < .05$.
\*\* $P < .01$.

### TABLE 4

MEAN OPERANT LEVEL AND MEAN AND ADJUSTED MEAN RESPONSES DURING LAST 15 MIN. OF CONDITIONING

| Group | Subgroup | N | Operant Level | | Last 15 Min. of Conditioning | | |
|-------|----------|---|------|------|------|------|------|
| | | | | | Obtained | | Adjusted |
| | | | Mean | SD | Mean | SD | Mean |
| H | Aware | 6 | 5.2 | 1.14 | 30.7 | 12.09 | 29.9 |
| | Unaware | 6 | 4.8 | 2.19 | 15.7 | 8.84 | 15.8 |
| | Control | 4 | 4.5 | 2.69 | 11.8 | 10.92 | 12.8 |
| FM | Aware | 7 | 7.4 | 1.99 | 53.1 | 14.50 | 55.1 |
| | Unaware | 5 | 7.8 | 2.71 | 28.0 | 18.86 | 28.2 |
| | Control | 4 | 8.8 | 3.77 | 31.0 | 20.23 | 26.8 |
| F | Aware | 3 | 14.3 | 7.41 | 48.7 | 17.99 | 44.9 |
| | Unaware | 9 | 8.4 | 5.87 | 44.2 | 14.02 | 45.6 |
| | Control | 4 | 10.2 | 2.48 | 40.2 | 11.09 | 40.0 |

humidity for each subgroup were 79.5–83.25° F. and 43.0–52.25%, respectively. There appeared to be no differences attributable to external environmental conditions over these slight ranges.

The success of this technique and the apparatus used was vouched for by all $S$s save one. In response to Question 2, in which $S$s were asked how well they liked the cool air, responses such as: "Very, very much"; "It was ecstasy"; "Desperately"; "Like a Manhattan"; "Couldn't have survived without it" were typical. In pilot studies in which $S$s were informed as to how they could obtain the cool air, some $S$s gave as many as eight responses per min. when six responses per min. were sufficient to maintain a constant supply of cool air. It was not uncommon for $S$s to ver-

bally request air, particularly during the extinction period, and $E$ suffered some verbal abuse from $S$s in all groups when these requests were not fulfilled.

## DISCUSSION

The results obtained in Groups H and FM require little interpretation. When plotted cumulatively, the extinction curves of the aware $S$s still show positive although reduced slope. This is not surprising since $S$s had been on a partial reinforcement schedule just prior to the extinction phase and the strength of the reward was great. Some confounding may also be present, since the controls in all three groups increased in response rate during extinction. Also to be noted is that operant levels did change as a part of a general change in activity during the period, exclusive of experimental manipulation, Fig. 1. The reason that $S$s in Group H were generally lower in response rate may be due to the fact that the pseudo task required those $S$s to be using their hands constantly.

The results of this study give added support to the increasing evidence against "learning without awareness." Only those $S$s who could verbalize a contingent relation between their behavior and the reinforcement demonstrated learning. Since learning did occur for these $S$s it would appear that the reinforcement was effective. Further, the fact that somewhat less than 50% of the $S$s became aware of the contingency would suggest that the task allowed room for "learning without awareness" had such a process been operative.

A finding that strikes us as quite remarkable was that over half of the $S$s did not learn. Results such as this are not surprising in studies utilizing verbal reinforcements where the "rewards" are secondary and have constituted a part of the general social background of $S$s for many years. However, in the present case the reward was primary and dramatic and gained the attention of all $S$s to such a degree that it seems surprising

that humans could spend 35 min. with the occurrence of this reward contingent upon their behavior and not learn the contingency.

A factor that may at least partially account for $S$s not learning or becoming aware of the contingency of the reward upon their behavior is the high operant level of the responses selected for conditioning. This is particularly true in Group F where only 3 $S$s became aware of the contingency. Since the reward lasted for 10 sec., most $S$s in this group would receive nearly continuous reinforcement by merely maintaining their operant level. In fact Table 4 demonstrates that the mean number of responses of the control $S$s in Group F was greater than that of the aware $S$s in Group H.

The data forcefully demonstrate the need for appropriate control $S$s in work on operant conditioning since it is obvious that parameters other than the experimentally controlled reinforcement exert an influence on $S$s' behavior. Also, the thorough exploration of $S$s' verbalized hypotheses and other behavior is shown to be an absolute requirement in determining awareness. Only 4 $S$s of 16 who were ultimately classified as aware volunteered the correct or correlated contingency in response to the initial question.

It appears that the technique utilized in this study offers a means to replicate on humans the behavioral changes found in subhuman animals. The high temperature and humidity, supplemented by appropriate instructions, allows for the manipulation of a reinforcement which definitely seems comparable in reward value to those administered to deprived animal $S$s. Isolation within the soundproof chamber also affords a more comparable situation, and one in which most stringent environmental and task controls are built in. Some limitations are placed on a complete survey of operant conditioning phenomena owing to the duration of the reward (determined in pilot work as the optimum), however this limitation appears to be slight when compared with the ease of administration and overall time saved by the procedure.

## SUMMARY

This study was concerned with a technique for producing operant conditioning of human motor responses with special emphasis on conditioning without awareness. Female $S$s who thought they were participating in a "space" experiment were enclosed in a chamber at 105° F. and relative humidity of 85%. A 10-sec. stream of cool air served as a reward for operant responses. Two of the three experimental groups (designated by response class) demonstrated conditioning and intensive investigation of $S$s' awareness revealed that conditioning was apparent only in those $S$s who were able to verbalize mediational steps that intervened between the experimental variables and the changes in their behavior.

## REFERENCES

AZRIN, N. H., HOLZ, W., ULRICH, R., & GOLDIAMOND, I. The control of the content of conversation through reinforcement. *J. exp. Anal. Behav.*, 1961, 4, 25–30.

BUCHWALD, A. M. Supplementary report: Alteration in the reinforcement value of a positive reinforcer. *J. exp. Psychol.*, 1960, 60, 416–417.

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, 67, 279–300.

HEFFERLINE, R. F., KEENAN, B., & HARFORD, R. A. Escape and avoidance conditioning in human subjects without their observation of the response. *Science*, 1959, 130, 1338–1339.

KRIECKHAUS, E. E., & ERIKSEN, C. W. A study of awareness and its effects on learning and generalization. *J. Pers.*, 1960, 28, 503–517.

LEVIN, S. M. The effects of awareness on verbal conditioning. *J. exp. Psychol.*, 1961, 61, 67–75.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, 51, 668–676.

VERPLANCK, W. S. The operant conditioning of human motor behavior. *Psychol. Bull.*, 1956, 53, 70–83.

# SIMULTANEOUS CONTRAST AS A FUNCTION OF TEST-FIELD AREA [1]

## A. LEONARD DIAMOND

*Psychological Research Center, University of Hawaii*

The phenomenon of simultaneous contrast is illustrated in the change in brightness of a visual field without a corresponding change in the luminance of that field. That is, if next to a small illuminated square, which we shall call the test field, we place another illuminated square of much greater luminance, which we shall term the inducing field, we will perceive a decrease in the brightness of the test-field square even though the test field is held constant in luminance.

A number of parameters of this phenomenon have been investigated in previous experiments. The luminance of the test and inducing fields (Diamond, 1953; Heinemann, 1955), the separation between test and inducing fields (Fry & Alpern, 1953; Leibowitz, Mote, & Thurlow, 1953), and the area of the inducing field (Diamond, 1955) have all been found to be pertinent variables in the contrast effect.

It is now of interest to know how the variation of the test-field area will affect simultaneous contrast. Specifically, a theoretical formulation (Diamond, 1960; also see Discussion below), which describes the relationships between the above-mentioned parameters, would necessarily predict little or no change in test-field brightness as the test-field area is varied in such a way that its center remains at a constant distance from the center of the inducing field. The present experiment is designed to test this prediction.

Our general method is as follows (see Fig. 1): To $S$'s right eye is presented an inducing field ($i$), a rectangle twice as wide as it is long; below the inducing is the test field ($t$), equally as wide as the inducing field but variable in height, and thus in area. To $S$'s left eye is presented the match field ($m$), which is either kept equal in size to the test field, as the test-field area is varied, or is held constant in size. (As we shall see, whether we vary the match-field size or hold it constant makes little difference in the results.) As the test-field area is varied, the distance between the centers of the test and inducing fields is held constant. The brightness of the test field is thus measured as a function of its area for three different values of test-field luminance (.69, 1.60, and 2.68 log mL.) over a wide range of inducing-field luminances ($-\infty$ to 2.79 log mL.).

## METHOD

### Apparatus

*Description of S's view.*—A modification of the apparatus used by Diamond (1955) is employed in the present experiment. The patterns seen by $S$ are different for each eye (see Fig. 1).

The pattern, R, to the right eye only, includes: (a) a test-field rectangle ($t$), 33' in visual angle along the horizontal, and variable in its vertical extent from zero to 33'; (b) above the test field with its center held at a constant distance from the center of the test field, an inducing-field rectangle ($i$), also 33' wide but with a fixed vertical dimension of 16.5'; and (c) a small fixation point (P) located 21' to the left of the test field. The

second pattern, L, presented to the left eye only, includes: (a) a match-field rectangle (m), 33′ and variable in height; and (b) a small fixation point (P) located 21′ to the right of the match field.

In the binocular view, S is instructed to fuse left- and right-eye fixation points into the point P so that the match, test, and inducing fields are held in constant position relative to one another. Stimulation may be considered essentially foveal since the visual angle between the fused fixation point and furthest corner of any field is never more than 70′.

In order to make it easier for S to fuse the left- and right-eye fixation points, a circular prism, the refracting angle of which is 5°, is placed in each eyepiece of the apparatus. By rotating each eyepiece, one clockwise and the other counterclockwise, S can optically rotate both left- and right-eye patterns such that he can vary the horizontal separation between them and thus more easily superimpose (fuse) the left- and right-eye fixation points. No systematic changes in the brightness of either right- or left-eye patterns occurred as a result of prism rotation.

*Apparatus controls.*—The apparatus is designed to control the following variables: (a) the luminances of the test, inducing, and match fields; and (b) the areas of the test and match fields. Luminance controls are discussed in detail by Diamond (1955). The general arrangement is as follows: The right- and left-eye patterns (L and R in Fig. 1) are presented to S along two separate optical paths, one to each eye. The light source of each path is a 150-w. tungsten filament projection lamp, in front of which is a section of heat-absorbing glass. The light, diffused by flashed opal glass, travels through its particular pattern (L or R). The luminance in either path can be continuously varied by fixed filters and a fixed and movable Polaroid. Made parallel by 4-diopter lenses, the light finally travels through 3-mm. artificial pupils into the eyes of S who is seated in a light-tight cubicle. The entire left optical path is adjustable horizontally for interpupillary distance by means of a screw arrangement attached to the optical bench.

The luminance of the match, test, and inducing fields are calibrated by means of binocular matches to fields of similar shapes and areas and whose luminances had been determined by a MacBeth illuminometer.

Test-field area is controlled by the use of six thin metal masks. In each of these masks is cut a test-field rectangle of a particular vertical dimension (5.5′, 11.0′, 16.5′, 22.0′, and 33.0′ in visual angle) and constant
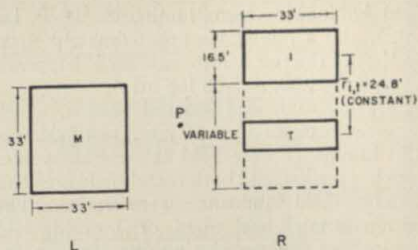
Fig. 1. The S's binocular view to left (L) and right (R) eyes. (The match field—m— was held either at its maximum of 33 min. or equal in area to the variable test-field—t.)

horizontal dimension (33.0′). An inducing-field rectangle of constant dimensions (16.5′ × 33.0′) is also cut in each mask such that its center is always at the same distance (24.75′) from the test-field center. This is illustrated as Pattern R in Fig. 1.

Match-field area is controlled as follows: A thin metal mask, in which is cut a 33′ square, is placed before the diffused light source (in the left optical path). Another thin piece of straight-edged metal is mounted immediately next to the square mask such that when the straight edge of this second piece of thin metal coincides with the top edge of the match field, the match-field area is zero. When the straight-edged metal is drawn down away from the top of the match field, the match-field area increases in successive rectangular increments. In this way, the match field can be set equal in area to any of the six test-field areas.

## Procedure

The experimental method was designed to investigate the brightness of the test field as a function of its area; the area was varied in six steps from 5.5′ to 33′ (vertical dimension). This entire function was studied for three test-field luminances, .69, 1.60, and 2.68 log mL., and for various inducing-field luminances as specified in Table 1.

It was desirable to determine whether variation of the match-field area would affect the results in any systematic way. Therefore, in addition to the variation of the aforementioned parameters, two conditions of match-field area were explored. That is, the match-field area was (a) held equal to that of the test field as the test-field area was varied and (b) held constant at its maximum value, i.e., 33′ square for all values of test-field area. This last condition prevailed for two inducing-

field luminances—zero luminance ($-\infty$ log mL.), and a luminance approximately equal to that of the test field, whatever it happened to be (.69, 1.60, or 2.68 log mL.).

During each experimental session, which required between 30 to 60 min., test-field area and luminance were held at constant values for that session, and both match-field area and inducing-field luminance were varied. The various match-field sizes or inducing-field luminances explored during one session were counterbalanced. Intersession periods were never shorter than 4 hr. Successive test-field areas and test-field luminances were taken in different sessions and were presented in random order.

*Psychophysical method.*—Seated in the light-tight cubicle, $S$ initially dark adapted for 3 min., then light adapted for 3 min. to the binocular view, i.e., at a test-field luminance and a set area for a particular experimental session, the match field at the same brightness as the test field, the inducing field at a luminance set for a particular experimental point. The $S$ then began making brightness-equality matches following the psychophysical method of adjustments as described by Guilford (1936). That is, $S$ set the luminance of the match field so that it appeared equal to

that of the test field. The $E$ then changed the luminance of the match field in a random manner after each match. The $S$ had to adjust the match-field luminance again until it seemed equal in brightness to that of the test field. In this manner, for each experimental point as indicated in Table 1, $S$ made 10 matches. The average of these 10 match-field luminances was taken as the brightness of the test field for each experimental point. This procedure was followed for 2 $S$s, JS and RH.

## Results

The data are presented for each $S$ in Table 1, and averaged in Fig. 2. Figure 2 shows the log luminance of the match field ($B_m$) plotted as a function of the test-field area (as measured by its vertical dimension). Log $B_m$ may be termed the brightness of the test field since its value is based upon an equality judgment between the match- and test-field brightnesses. The function is graphed for different values of test-field luminances (log $B_t$),

TABLE 1

Log Match-Field Luminances (mL.) for Different Test Area
and Inducing and Test Luminances

| Log Test Luminance (mL.) | Log Ind. Luminance (mL.) | Test Height (Min.) | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | 5.5 | | 11.0 | | 16.5 | | 22.0 | | 27.5 | | 33.0 | |
| | | RH | JS | RH | JS | RH | JS | RH | JS | RH | JS | RH | JS |
| 0.69 | $-\infty$ | .73 | .86 | .46 | .75 | .57 | .82 | .50 | .75 | .51 | .92 | .69 | .80 |
| | $-\infty_m$ | .64 | .99 | .52 | .77 | .37 | .88 | .48 | .66 | .53 | .81 | .68 | .81 |
| | 0.69 | .61 | .65 | .64 | .80 | .73 | .90 | .48 | .62 | .41 | .53 | .61 | .75 |
| | $0.70_m$ | .60 | .79 | .59 | .84 | .41 | .85 | .50 | .69 | .50 | .66 | .62 | .74 |
| | 1.70 | .26 | .50 | .46 | .59 | .29 | .34 | .34 | .44 | .09 | .36 | .11 | .49 |
| | 2.68 | $-.02$ | .06 | .23 | .30 | .09 | .11 | $-.07$ | .19 | $-.26$ | $-.03$ | $-.04$ | .22 |
| 1.60 | $-\infty$ | 1.28 | 1.91 | 1.31 | 1.76 | 1.49 | 1.81 | 1.49 | 1.66 | 1.40 | 1.73 | 1.40 | 1.58 |
| | $-\infty_m$ | 1.75 | 2.01 | 1.47 | 1.79 | 1.28 | 1.70 | 1.37 | 1.54 | 1.51 | 1.85 | 1.50 | 1.65 |
| | 1.60 | 1.31 | 1.66 | 1.21 | 1.54 | 1.49 | 1.69 | 1.32 | 1.47 | 1.24 | 1.61 | 1.53 | 1.58 |
| | $1.69_m$ | 1.42 | 1.96 | 1.33 | 1.72 | 1.22 | 1.56 | 1.38 | 1.54 | 1.33 | 1.45 | 1.17 | 1.49 |
| | 2.23 | 1.33 | 1.70 | 1.25 | 1.46 | 1.18 | 1.38 | 1.18 | 1.37 | 1.39 | 1.53 | 1.26 | 1.75 |
| | 2.72 | 1.02 | 1.31 | .99 | 1.17 | 1.01 | 1.30 | 1.18 | 1.14 | 1.02 | 1.11 | 1.26 | 1.45 |
| 2.68 | $-\infty$ | 2.19 | 2.84 | 2.48 | 2.61 | 2.37 | 2.72 | 2.51 | 2.83 | 2.50 | 2.86 | 2.73 | 2.88 |
| | $-\infty_m$ | 2.59 | 2.99 | 2.80 | 2.94 | 2.46 | 2.76 | 2.45 | 2.81 | 2.51 | 2.85 | 2.65 | 2.92 |
| | $-0.32$ | 2.20 | 2.74 | 2.60 | 2.50 | 2.08 | 2.74 | 2.29 | 2.77 | 2.48 | 2.66 | 2.65 | 2.76 |
| | 1.22 | 2.35 | 2.83 | 2.50 | 2.87 | 2.31 | 2.62 | 2.53 | 2.82 | 2.36 | 2.68 | 2.60 | 2.82 |
| | 2.68 | 2.30 | 2.75 | 2.47 | 2.62 | 2.38 | 2.47 | 2.49 | 2.72 | 2.27 | 2.61 | 2.66 | 2.70 |
| | $2.79_m$ | 2.64 | 2.94 | 2.42 | 2.73 | 2.21 | 2.56 | 2.59 | 2.80 | 2.40 | 2.71 | 2.71 | 2.71 |

* The subscript $m$ indicates a constant maximum match-field area for all test areas.
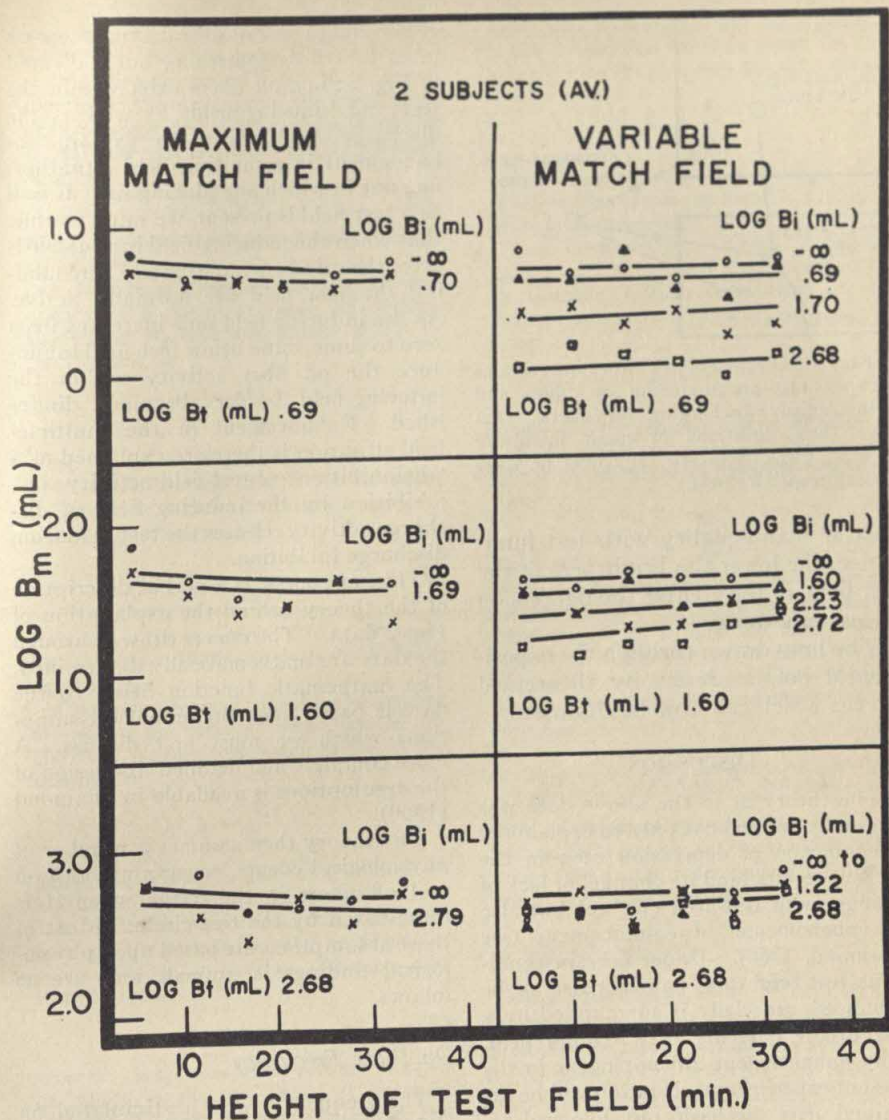
FIG. 2.  Test brightness as a function of test area for different test
and inducing luminances.

and for different inducing-field lumin-
ances (log $B_i$) and match-field areas.

The effect of test-field area upon
test-field brightness is minimal.
There seems to be little or no change
in the brightness of the test field as its
area is increased; this holds through-

out the test-field and inducing-field
luminance range explored and for both
conditions of match-field area. In-
ducing luminance, however, does have
an effect as seen in previous experi-
ments (Diamond, 1953, 1955).  The
greater the inducing luminance,
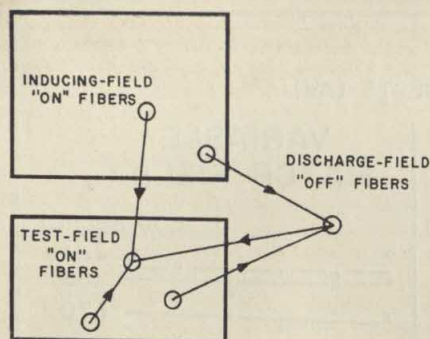
FIG. 3. Diagram of inhibitory interactions between test and inducing on fibers and spontaneously discharging off fibers. (Arrows show directions of main inhibitory effects. Circles represent any fiber or fibers in each respective field.)

greater than equality with test luminance, the lower the brightness of the test field, or the greater the depression of test brightness.

The lines drawn through the experimental points are fit by theoretical curves which can now be discussed.

## DISCUSSION

The theory fit to the above data was originally devised not only to explain the phenomenon of depression seen in the present results (and its change, or lack of change, with test-area change) but also the phenomenon of enhancement (see Diamond, 1960). Under certain conditions test brightness can increase, or be enhanced, especially if surrounded by a less bright inducing or surround field. Why enhancement did not occur in the present experimental situation will be explained after the basic physiological assumptions to the theory are summarized. In Fig. 3 are diagramed the main or primary physiological events that according to our theory are basic to psychophysical depression and enhancement.

For depression of test brightness, the test field "on" fibers (which presumably mediate psychophysical brightness) are inhibited by (a) the inducing on fibers, (b) other test on fibers, and (c) the spontaneously discharging off fibers.

For enhancement of test brightness we must first note something not indicated in Fig. 3; i.e., off fibers exist within the test and inducing fields as well as the discharge field. Now to explain enhancement in a multiple-field situation; i.e., one in which an inducing field as well as a test field is present, we must assume that when the inducing-field luminance is zero, the off fibers immediately surrounding the test field are normally active. As the inducing field now increases from zero to some value below test-field luminance the off fiber activity within the inducing-field borders becomes diminished. Enhancement in the multiple-field situation is therefore explained as a "disinhibition" of test-field activity; i.e., inhibition by the inducing field of discharge activity releases the test field from discharge inhibition.

This, of course, is a verbal description of the theory behind the explanation of Fig. 2 data. The curves drawn through the data are mathematically determined. The mathematic function fitted to the data is based upon preliminary assumptions which we must first discuss. A more complete and detailed discussion of the assumptions is available in Diamond (1960).

Our theory then assumes a number of physiological events, concerning both on and off fibers in the retina when it is illuminated by the test circle. Most of these assumptions are based upon physiological findings in animals and are as follows.

### On Fiber Frequency

The frequency $f_{t_1}$ of a stimulated on retinal fiber (as described by Hartline, 1938):

1. directly determines the strength of the brightness response, or

$$A_t = k_1 f_{t_1} \qquad [1]$$

where $A_t$ represents the brightness response, $k_1$ is a proportionality constant, and $f_{t_1}$ is the frequency of the on fiber. This relationship has been suggested by Adrian's (1928) demonstration of the similarity between the brightness-duration curves in

man and the frequency-duration curves taken from the optic nerve of the eel.

2. is directly proportional to a power function of the luminance of the stimulating light, as suggested in Diamond's (1960) fit of Hartline and Graham's (1932) *Limulus* data, or

$$f_{t_1} = k_2 B_{t_1}^{\frac{1}{2}} \qquad [2]$$

where $B_{t_2}$ is the luminance of light striking an on fiber.

3. is inversely proportional to the frequency of a nearby on fiber, as demonstrated by Hartline and Ratliff (1957) in the *Limulus*, or

$$f_{t_1} = \frac{k_3}{f_{t_2}} \qquad [3]$$

where $f_{t_2}$ is the frequency of a nearby on fiber within the test field. Since according to Equation 2 above frequency is proportional to luminance then

$$f_{t_2} = \frac{k_4}{B_{t_2}^{\frac{1}{2}}} \qquad [4]$$

when $B_{t_2}$ is the luminance striking the nearby on fiber. Within the test field this is equal to $B_{t_1}$.

4. is inversely proportional to the number of nearby on fibers as demonstrated by Hartline, Wagner, and Ratliff (1956) in the *Limulus*, or

$$f_{t_1} = \frac{k_5}{E_{t_2}} \qquad [5]$$

where $E_{t_2}$ represents the total number of nearby on fibers within the test field.

5. is directly proportional to the distance between the test and nearby on fibers, as demonstrated by Hartline, Wagner, and Ratliff (1956), or

$$f_{t_1} = k_6 r_{t,t} \qquad [6]$$

where $r_{t,t}$ is the separation between the two on fibers in the test field. It becomes convenient and actually desirable, as pointed out by Diamond (1955), to combine Equations 5 and 6 such that

$$C_{t,t} = \frac{k_7 E_{t_2}}{\bar{r}_{t,t}} \qquad [7]$$

where $\bar{r}_{t,t}$ is the average separation between all test fibers and $C_{t,t}$ therefore describes the combined effect of all the individual nearby on fibers; it then follows that

$$f_{t_1} = \frac{k_6 \bar{r}_{t,t}}{E_{t_2}} = \frac{k_8}{C_{t,t}} \qquad [8]$$

Equations 2 through 8 are also applicable to the interaction between a test on fiber and inducing on fiber (see Fig. 3). Thus

$$f_{t_1} = \frac{k_9}{C_{i,t}} \qquad [9]$$

where $C_{i,t}$ represents the combined effect of all on fibers from the inducing field upon those in the test field.

6. is inversely proportional to the frequency of a nearby off fiber, as suggested in Granit's (1955) descriptions of the mutual antagonism between on and off fibers, or

$$f_{t_1} = \frac{k_{10}}{f_o} \qquad [10]$$

where $f_o$ is the frequency of an off fiber.

7. is inversely proportional to the number of nearby off fibers at particular distances away or

$$f_{t_1} = \frac{k_{11}}{C_{d,t}} \qquad [11]$$

where $C_{d,t}$ is the combined effect of all the off fibers in the discharge field on the on fiber frequency in the test field. With respect to the frequency of an on retinal fiber in the test field, Equations 1 through 11 may be combined into the following formula:

$$f_t = \frac{K_o B_t^{\frac{1}{2}}}{K_1 B_t^{\frac{1}{2}} K_2 C_{t,t} + K_3 B_t^{\frac{1}{2}} K_4 C_{i,t} + K_5 f_o C_{d,t}} \qquad [12]$$

The subscripts of the proportionality constants $(K)$ are chosen to coincide with those used in the more general brightness theory by Diamond (1960).

## Off Fiber Frequency

The frequency, $f_o$, of an off fiber in the discharge field, which we assume to be spontaneously active in the dark according to data taken from the cat eye by Barlow, FitzHugh, and Kuffler (1954):

1. is directly proportional to some "internal driving force" (comparable to the external luminance effect on on fibers) which we shall assume to exist. Such a mechanism is suggested by experiments, described by Granit (1955), which show both on and off activity to increase in the retina as a result of central (reticular formation) stimulation. Thus

$$f_o = k_{13} D^{\frac{1}{2}} \qquad [13]$$

where $D$ represents this "driving force."

2. is inversely proportional to the frequency of a nearby on test fiber, according to Granit's on-off antagonism findings cited above, or

$$f_o = \frac{k_{14}}{f_t} \quad [14]$$

3. is inversely proportional to number and distances away of nearby on test fibers, or

$$f_o = \frac{k_{15}}{C_{t,d}} \quad [15]$$

where $C_{t,d}$ is the combined effect of all the on fibers in the test field upon the off fibers in the discharge field. This assumption has not been tested experimentally.

4. is inversely proportional to the amount of light impinging upon the off fiber, as demonstrated by Hartline (1938). This light could be direct or scattered from the test beam stimulating the on fibers. The retinal effectiveness of scattered light in the human eye has been demonstrated by Boynton and Riggs (1951). Thus,

$$f_o = \frac{k_{16}}{B_t} \quad [16]$$

With respect to the frequency of an off fiber in the retinal discharge field, therefore, Equations 12 through 15 may be combined into the following formula:

$$f_o = \frac{K_5 D^{\frac{1}{2}}}{K_7 B_t C_{t,d}} \quad [17]$$

Equations 14 through 16 are also applicable to the interaction between an inducing on fiber and a discharge off fiber (see Fig. 3). Thus

$$f_o = \frac{K_5 D^{\frac{1}{2}}}{K_8 B_i C_{t,d}} \quad [18]$$

where $C_{i,d}$ represents the combined effect of all on fibers from the inducing field upon the off fibers in the discharge field.

It should be noted that the inhibition of the discharge field by the in-

ducing field is effectively greater when an inducing square, for example, is near the test square. This is because discharge elements near the test field are initially more effective than those far away. When the inducing field inhibits near discharge elements, the effect of spontaneous discharge on the test field is much more reduced than when the far discharge elements are inhibited. Therefore the effective discharge frequency is a function of the separation between test and inducing fields, or

$$f_o = k_{17} \frac{\bar{r}_{i,t}}{\bar{r}_{t,t}} \quad [19]$$

where $\bar{r}_{i,t}$ is the average separation between all elements in the inducing field and those in the test field. That $\bar{r}_{i,t}$ is set in ratio to $\bar{r}_{t,t}$ is empirically required for a satisfactory fit of certain data (see Diamond, 1960, pp. 183–184). A completely rational account of the functional relationship between $f_o$ and $\bar{r}_{i,t}$ awaits further knowledge of separate and relative effects of light scatter in the intact human eye and on-off antagonism.

Equations 17 through 19 may be combined to include the effect of the inducing field upon the off frequency, or

$$f_o = \frac{K_5 D}{K_7 B_t^{\frac{1}{2}} C_{t,d} + K_8 B_i^{\frac{1}{2}} \dfrac{\bar{r}_{t,t} C_{i,d}}{\bar{r}_{i,t}}} \quad [20]$$

If we now combine Equations 1, 12, and 20 then

$$A_t = \frac{K_o B_t^{\frac{1}{2}}}{K_1 B_t^{\frac{1}{2}}(K_2 C_{t,t}) + K_3 B_t^{\frac{1}{2}} K_4 C_{i,t} + \dfrac{K_5 D^{\frac{1}{2}}(K_6 C_{d,t})}{K_7 B_t C_{t,d} + K_8 B_i \dfrac{\bar{r}_{t,t} C_{i,d}}{\bar{r}_{i,t}}}} \quad [21]$$

The $S$'s brightness response to the match field can be described in the same manner minus the inducing field effect, or

$$A_m = \frac{K_o B_m^{\frac{1}{2}}}{K_1 B_m^{\frac{1}{2}}(K_2 C_{m,m}) + \dfrac{K_5 D^{\frac{1}{2}}(K_6 C_{d,m})}{K_7 B_t C_{t,d}}} \quad [22]$$

Since in the present experiment, the $S$ is required to match the test and match field in brightness such that

$$A_t = A_m \quad [23]$$

then if accordingly we set Equations 21 and 22 equal

$$\frac{K_o B_m^{a/4}}{K_1 B_m^{b_1/4}(K_2 C_{m,m})^{c_1} + \dfrac{K_5 D^{b_4/4}(K_6 C_{d,m})^{c_2}}{K_7 B_m C^e_{m,d}}}$$

$$= \frac{K_o B_t^{a/4}}{K_1 B_t^{b_2/4}(K_2 C_{t,t})^{c_3} + K_3 B_i^{b_3/4}(K_4 C_{i,t})^{c_4} + \dfrac{K_5 D^{b_4/4}(K_6 C_{d,t})^{c_5}}{K_7 B_t C^e_{t,d} + K_8 B_i \left(\dfrac{\bar{r}_{t,t} C_{i,d}}{\bar{r}_{i,t}}\right)^f}} \quad [24]$$

In Table 2 are presented the values of the terms in Equation 24. The procedures for determining values for the terms were as follows: The values of the luminance ($B$) terms were taken directly from the luminance values of the different fields. The area-separation ($C$) terms were calculated according to Equation 10 above. The dimensions of the discharge area were arbitrarily chosen; since only minimal inhibitory effects occur between fields separated by more than 4.5° (Fry & Alpern, 1953) of visual angle and at certainly no more than 9° (Leibowitz et al., 1953); spontaneous discharge elements further than 9° from the center of the test field were not considered to be effective.

The value of $D$ (spontaneous discharge activity) is arbitrarily taken as the value 1. This turns out to be empirically satisfactory and must do until such time as physiological and/or psychophysical, manipulation of discharge activity reveals its actual value.

The value of the $a$, $b$, $c$, $e$, $f$, and $K$ terms were determined empirically

(by trial and error fitting). The value of $K_o$ is not included in Table 2 since this term cancels out in Equation 20.

It should be noted that in Equation 20 and Table 2 are listed a total of 18 fitted constants. Were these constants chosen completely without restriction, any number of a variety of functions could be generated, so that the fit of our physiological theory to the 16 functions in our present experiment (see Fig. 2) would be meaningless. The constants, however, were not chosen arbitrarily but were those fit by Diamond (1960) to more than 45 additional functions having to do with brightness contrast and predicted by the same theory as outlined above.

As noted above, enhancement as a function of test area is not clearly evident in the data. Furthermore, although it is included as a basic component of the present theory it is not predicted for the stimulus conditions existing in the present experiment. That is, the off fibers in the discharge field (see Fig. 3) are already inhibited

## TABLE 2
### VALUES OF TERMS IN EQUATION 24

| | Determined by Experimental Situation | | | | | | | | | Empirically Fitted Values | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Luminance (Log mL) | | | Area-Separation (Log) | | | Exponents | | | Proportionality Constants (Log) | | | "Off Drive" | | |
| | Terms | Max. Match | Variable Match | Terms | Max. Match | Variable Match | Terms | Max. Match | Variable Match | Terms | Max. Match | Variable Match | Terms | Max. Match | Variable Match |
| | $B_m$ | v | v | $C_{m,m}$ | 2.62 | v | $a$ | 12.00 | 12.00 | $K_1$ | 0.00 | 0.00 | $D$ | 1.00 | 1.00 |
| | $B_t$ | v | v | $C_{d,m}$ | 4.19 | 4.19 | $b_1$ | 4.00 | 4.00 | $K_2$ | −2.80 | −2.80 | | | |
| | $B_i$ | v | v | $C_{m,d}$ | 1.28 | v | $b_2$ | 4.00 | 4.00 | $K_3$ | −0.00 | −0.00 | | | |
| | | | | $C_{t,t}$ | v | v | $b_3$ | 4.00 | 4.00 | $K_4$ | −2.28 | −2.28 | | | |
| | | | | $C_{i,t}$ | 2.14 | 2.14 | $b_4$ | 10.92 | 10.92 | $K_5$ | 0.30 | 0.30 | | | |
| | | | | $C_{d,t}$ | 4.19 | 4.19 | $c_1$ | 0.30 | 0.30 | $K_6$ | −3.75 | −3.75 | | | |
| | | | | $C_{t,d}$ | v | v | $c_2$ | v | v | $K_7$ | 0.00 | 0.00 | | | |
| | | | | $C_{d,d}$ | 1.28 | 1.28 | $c_3$ | 0.30 | 0.30 | $K_8$ | −3.40 | −3.40 | | | |
| | | | | $\hat{r}_{t,t}$ | v | v | $c_4$ | v | v | | | | | | |
| | | | | $\hat{r}_{i,t}$ | | | $c_5$ | v | v | | | | | | |
| | | | | | | | $e$ | 2.00 | 2.00 | | | | | | |
| | | | | | | | $f$ | 3.00 | 3.00 | | | | | | |

Note.—v = variable.

$c = \left(\frac{b}{4}\right)^{1.1} \log\left(\frac{B_i}{B_t} + 1\right)$.   This is empirically derived and not yet based upon theory.

by the suprathreshold test light, according to Equation 17 above. The increase in test area therefore has no further effect on the discharge off fibers thereby resulting in little or no disinhibition or enhancement.

## SUMMARY

The brightness of a test-field rectangle was studied as a function of its area and the luminance of a nearby inducing-field rectangle. The area of the test field was varied by increasing its vertical dimension from 5.5' to 33.0' in visual angle. A binocular matching technique was used in which $S$ indicated the test-field brightness by the luminance of a match-field set equal in brightness to the test field.

Two experiments were performed. In the first the match area was held at a constant maximum value. In the second the match area was allowed to vary along with the test area. The dependent variable for both experiments was match-field luminance and the independent variables, test area, test luminance, and inducing luminance. Very little if any change occurred in the test-field brightness as a function of test-area variation in either experiments. This held true for a wide range of test and inducing luminances. As inducing luminance was increased, however, test brightness became depressed. A mathematical theory fit to the data was based upon physiological inhibitive interaction among individual cone receptors in the retina.

## REFERENCES

ADRIAN, E. D. *The basis of sensation.* London: Christopher, 1928.

BARLOW, H. B., FITZHUGH, R., & KUFFLER, S. W. Resting discharge and dark adaptation in the cat. *J. Physiol.*, 1954, **125**, 28–29.

BOYNTON, R. M., & RIGGS, L. A. The effect of stimulus area and intensity upon the human retinal response. *J. exp. Psychol.*, 1951, **42**, 217–226.

DIAMOND, A. L. Foveal simultaneous brightness contrast as a function of inducing and test-field luminances. *J. exp. Psychol.*, 1953, **43**, 189–195.

DIAMOND, A. L. Foveal simultaneous contrast as a function of inducing-field area. *J. exp. Psychol.*, 1955, **50**, 144–152.

DIAMOND, A. L. A theory of depression and enhancement in the brightness response. *Psychol. Rev.*, 1960, **67**, 168–198.

FRY, G. A., & ALPERN, M. The effect of a peripheral glare source upon the apparent brightness of an object. *J. Opt. Soc. Amer.*, 1953, **43**, 189–195.

GRANIT, R. *Receptors and sensory perception.* New Haven: Yale Univer. Press, 1955.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

HARTLINE, H. K. The response of single optic nerve fibers of the vertebrate retina. *Amer. J. Physiol.*, 1938, **121**, 400–415.

HARTLINE, H. K., & GRAHAM, C. H. Nerve impulses from single receptors in the eye. *J. cell. comp. Physiol.*, 1932, **1**, 277–295.

HARTLINE, H. K., & RATLIFF, F. Inhibitory interaction of receptor units in the eye of *Limulus*. *J. gen. Physiol.*, 1957, **40**, 375–376.

HARTLINE, H. K., WAGNER, H. G., & RATLIFF, F. Inhibition in the eye of *Limulus*. *J. gen. Physiol.*, 1956, **39**, 651.

HEINEMANN, E. G. Simultaneous brightness induction as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1955, **50**, 89–96.

LEIBOWITZ, H., MOTE, F. A., & THURLOW, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *J. exp. Psychol.*, 1953, **46**, 454–456.

# MEDIATED SATIATION IN VERBAL TRANSFER [1]

## LEON A. JAKOBOVITS AND WALLACE E. LAMBERT

### *McGill University*

The role of mediation in associative processes has long been recognized by psychologists (see Atherton & Washburn, 1912) despite early failures to demonstrate its influence in controlled experiments (e.g., Howe, 1893). Peters (1935) was among the first to report positive results in controlled designs, and recently, several investigators (McGehee & Schulz, 1962; Russell & Storms, 1955) showed conclusively that mediation in paired-associate learning can be demonstrated in the laboratory. In these studies the existence of mediation in verbal learning was inferred from transfer designs permitting positive generalization of acquisition gradients. The present paper represents an attempt to study mediation in verbal transfer using generalization of inhibition gradients. In other words, a design was used which, on the basis of previously established facts, should lead to facilitative effects in paired-associate learning but, because the inferred mediators are "tampered with," inhibition instead of facilitation is expected to be transferred. Since the present design produces inhibition by the direct manipulation of the mediators, it has an advantage over previous approaches in which the mediated effect remained at the implicit level.

## METHOD

*Design.*—The design of the present study was parallel to those of Russell and Storms (1955) and McGehee and Schulz (1962). In Group E (experimental), Ss learned two paired-associate lists: List 1 established A-B connections between nonsense syllables and meaningful words; List 2 was composed of A-D pairs, where the relation between B and D was such that D was the most common associate to C and the latter was the most common associate to B. The middle element, C, acts as the mediating link in the forward associative chain, $B \rightarrow C \rightarrow D$, and provides facilitation in the acquisition of the A-D list. In the present case, however, the C word was "satiated" according to a technique described by Lambert and Jakobovits (1960) who showed that continuous repetition of a word results in a decrement in the intensity of its connotative meaning. It was expected that the decreased meaningfulness of the C word would reduce its effectiveness to act as a mediator in the $B \rightarrow C \rightarrow D$ chain, reducing the facilitation effect during the subsequent acquisition of the A-D pairs.

In Group C (control), Ss also learned two lists: List 1 established A-X associations between the same nonsense syllables but other meaningful words than those used for Group E; List 2 was composed of the same A-D pairs as used for Group E. However, no associative relation existed between the X and D words. The words were the same as those used in the Russell and Storms (1955) study where a complete description of the procedure for selecting words can be found.

The overall design is illustrated in Table 1. Group E received the Mediator Nonsatiated and the Mediator Satiated conditions. Each of these two conditions consisted of five paired associates. The $A_1$-$B_N$ and $A_1$-$D_N$ pairs correspond to the second half of the A-B and A-D pairs, respectively, in Table 3 of the Russell and Storms study. The $A_2$-$B_S$ and $A_2$-$D_S$ pairs correspond to the first half of the A-B and A-D pairs, respectively, in their study. The 10 A-B pairs formed List 1 in the present study, while the 10 A-D pairs formed List 2. Three different random orders of the 10 pairs

## TABLE 1

### ILLUSTRATION OF THE DESIGN USED IN THE EXPERIMENT

| Condition | List 1 | List 2 | Inferred Action |
|---|---|---|---|
| Mediator Nonsatiated | $A_1$-$B_N$ | $A_1$-$D_N$ | $B_N \longrightarrow C_N$ <br> $A_1 \cdots\cdots\cdots \longrightarrow D_N$ |
| Mediator Satiated | $A_2$-$B_S$ | $A_2$-$D_S$ | $B_S \longrightarrow C_S$ <br> $A_2 \cdots\cdots\cdots \longrightarrow D_S$ |
| Nonmediated Control | $A_1$-$X_1$ | $A_1$-$D_N$ | $X_1 \longrightarrow ?$ <br> $A_1 \cdots\cdots\cdots \longrightarrow D_N$ |
| | $A_2$-$X_2$ | $A_2$-$D_S$ | $X_2 \longrightarrow ?$ <br> $A_2 \cdots\cdots\cdots \longrightarrow D_S$ |

### Examples of the Above Conditions in the Same Order

| List 1 | List 2 | Inferred Action |
|---|---|---|
| GEX-JUSTICE | GEX-WAR | JUSTICE $\longrightarrow$ PEACE <br> GEX $\cdots\cdots\cdots \longrightarrow$ WAR |
| YOV-SOLDIER | YOV-NAVY | SOLDIER $\longrightarrow$ ARMY <br> YOV $\cdots\cdots\cdots \longrightarrow$ NAVY |
| GEX-HOUSE | GEX-WAR | HOUSE $\longrightarrow$ ? <br> GEX $\cdots\cdots\cdots \longrightarrow$ WAR |
| YOV-CHEESE | YOV-NAVY | CHEESE $\longrightarrow$ ? <br> YOV $\cdots\cdots\cdots \longrightarrow$ NAVY |

Note.—The notations $A_1$ and $A_2$ refer, respectively, to the first 5 and second 5 stimulus members of the 10-item lists. Similarly, $B_N$ or $D_N$ and $B_S$ or $D_S$ refer to the first and second half of the response members.

in each list were presented in a standard memory drum at a 3:3-sec. rate with a 6-sec. intertrial interval. The instructions given were the same as those described in detail by Storms (1958). Each $S$ saw List 1 for a maximum of 27 trials or until he met the criterion of three errorless repetitions, whichever came first. (All $S$s met the lesser criterion of one errorless repetition, but 10 failed to meet the criterion of three errorless repetitions within the maximum of 27 trials.) Eight minutes elapsed between the last presentation of List 1 and the first presentation of List 2. During this period, $S$s of Group E received the satiation treatment which involved the five $C_S$ words of the Mediator Satiated condition ($A_2$-$B_S$), as well as five filler words (actually, the second half of the X words in the Russell and Storms table). First, $S$ rated the 10 words (randomly mixed)

## TABLE 2

### NUMBER OF TRIALS TO CRITERION ON THE VARIOUS CONDITIONS FOR GROUPS E AND C

| Condition | Group E | | Group C | | $t$ between Groups |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | |
| $A_1$-$B_N$ and $A_2$-$B_S$ (List 1) | 19.80[a] | 5.09 | — | — | 1.26 |
| $A_1$-$X_1$ and $A_2$-$X_2$ (List 1) | — | — | 17.95[b] | 4.22 | |
| $A_1$-$D_N$ and $A_2$-$D_S$ (List 2) | 10.77[a] | 4.85 | 17.30[b] | 5.16 | 4.45** |
| $A_1$-$D_N$ | 7.77 | 3.45 | 15.25 | 5.14 | 6.13** |
| $A_2$-$D_S$ | 10.30 | 4.74 | 14.45 | 5.16 | 2.86* |

[a] The difference between these two means is significant ($t = 11.73$; $P < .001$).
[b] The difference between these two means is not significant ($t = 0.64$; $ns$).
* $P < .01$.
** $P < .001$.

on three scales of the semantic differential (pleasant-unpleasant, strong-weak, and fast-slow). Then $S$ repeated aloud each word for 20 sec. before rating it again on the same three scales. Differences in intensity of ratings before and after repetitions represent the semantic satiation scores to be presented below. Following the satiation treatment, List 2 was presented and $S$ learned it to a criterion of three errorless repetitions and in this case all $S$s reached the criterion in less than 27 trials.

The $S$s in Group C received the Nonmediated Control condition shown in Table 1. The five $A_1$-$X_1$ pairs and the five $A_2$-$X_2$ pairs formed List 1 which corresponds to the A-X column of the Russell and Storms table. The $A_1$-$D_N$ and $A_2$-$D_S$ pairs which formed List 2 were the same as those for Group E. The procedure used with Group E was duplicated here except for the fact that no satiation treatment was administered, and $S$s were engaged in neutral conversation during the 8 min. which separated the two lists.

In summary, the following comparisons can be noted: the Mediator Nonsatiated condition in the present study corresponds to the "Chained" condition of Russell and Storms (1955) or the "Mediated" group of McGehee and Schulz (1962). The Non-

mediated Control condition corresponds to the "Unchained" or "Nonmediated" conditions respectively in those two studies; the Mediator Satiated condition represents the proactive interference design of the present study. The results are presented in terms of differences in acquisition scores of List 2 under the three conditions.

*Subjects.*—The $S$s were 50 male English-speaking cadets of the Royal Canadian Air Force enrolled in a 6-wk. training course at a base near Montreal. They were asked to volunteer for the experiment by their instructor. The testing was done individually at the training base during regular work hours. The first 30 $S$s formed Group E, the last 20, Group C.

## RESULTS

Comparison between the two groups on the number of trials required to reach the criterion of one errorless repetition (met by all $S$s) for their particular first list is presented at the top of Table 2. There is no significant difference between the two groups, indicating that they are of

TABLE 3

COMPARISONS BETWEEN THE MEDIATOR NONSATIATED ($A_1$-$D_N$) AND MEDIATOR
SATIATED ($A_1$-$D_S$) CONDITIONS OF LIST 2 FOR GROUPS E AND C

| Measures | Means for Group E | | | Means for Group C | | |
|---|---|---|---|---|---|---|
| | $A_1$-$D_N$ | $A_2$-$D_S$ | Diff. | $A_1$-$D_N$ | $A_2$-$D_S$ | Diff. |
| First 5 diff. correct responses | 3.13 (0.62) | 1.87 (0.62) | 1.26** (1.72) | 2.45 (1.07) | 2.55 (1.07) | −0.10 (2.18) |
| Trials to 1 errorless repet. | 7.77 (3.47) | 10.30 (4.73) | −2.53** (3.28) | 15.25 (5.14) | 14.45 (5.15) | 0.80 (4.05) |
| Total correct anticipations[a] | 33.07 (15.40) | 26.70 (11.83) | 6.37** (6.41) | 43.30 (17.10) | 44.10 (15.50) | −0.80 (14.52) |
| Total number of intrusions[a] | 3.97 (3.43) | 6.30 (4.60) | −2.33* (3.50) | 12.30 (6.45) | 12.80 (7.65) | −0.50 (8.07) |
| Total number of omissions[a] | 17.67 (11.43) | 21.71 (13.83) | −4.04** (4.78) | 30.25 (7.55) | 28.95 (8.85) | 1.30 (7.98) |
| Total correct anticipations[b] | 47.90 (16.20) | 40.83 (11.83) | 7.07** (7.37) | 56.10 (14.85) | 57.25 (13.75) | −1.15 (13.19) |

Note.—Numbers in parenthesis are $SD$s.
[a] To a criterion of one errorless repetition.
[b] To a criterion of three errorless repetitions or 27 trials.
* $P < .01$, two-tailed $t$ test.
** $P < .001$, two-tailed $t$ test.

equal learning ability. This conclusion is correct unless the materials in the two lists are not of equal difficulty which is unlikely in view of the particular selection procedure used by Russell and Storms (1955) and the fact that McGehee and Schulz (1962), using the same meaningful words but other nonsense syllables of comparable association values, found no such difference. The two groups differ significantly on the acquisition of List 2, with Group E showing a marked superiority over Group C. This finding is a replication of McGehee and Schulz and other studies in which the mediated condition was found to be superior to a nonmediated condition. The same finding is pointed up by the fact that the acquisition of List 2 by Group E is significantly faster than for List 1, whereas no such facilitation effect is noticeable for Group C (see Footnotes a and b to Table 2). Breaking down the analysis of List 2 acquisition into $A_1$-$D_N$ and $A_2$-$D_S$ pairs it can be seen that in both cases Group E is significantly superior to the control. This means that the predicted proactive inhibition effect of the Mediator Satiated condition ($A_2$-$D_S$) did not result in absolute negative transfer. In fact, facilitation was noticed, although significantly less than in the case of the Mediator Nonsatiated condition ($A_2$-$D_N$), as will be indicated below.

The mean semantic satiation score for Group E on all 10 words was $-3.07$ ($SD = 6.64$; $t = 2.49$; $P < .02$). For the five $C_S$ words the mean was $-2.27$ ($SD = 3.82$; $t = 3.20$; $P < .01$); the mean for the five filler words was $-0.80$ ($SD = 4.09$; $t = 1.05$; $ns$). This difference in amount of semantic satiation shown on the two sets of words approaches significance ($t = 1.84$; $P < .10$), even though the product-moment correlation coefficient

between the two sets is significant ($r = .43$; $P < .02$). Although there is a basis for speculating on the reason for the difference noted (e.g., the $C_S$ words are related to the B words seen in List 1, whereas the filler words are not so related), such a discussion is not directly relevant to the purpose of this study.

Let us then turn to a comparison of the acquisition scores between the $A_1$-$D_N$ and $A_2$-$D_S$ pairs of List 2 for Group E. Russell and Storms (1955), making a similar comparison in their study, used two separate criteria. One of these was an analysis of the first five different correct responses made by the $S$s, with a view to determine whether the response terms of the A-D pairs, for which associative chaining was possible (here, the $A_1$-$D_N$ pairs), were more easily elicited during the early trials. In the present study these were compared to the $A_2$-$D_S$ pairs where the mediator was satiated. The other analysis involved subtracting the total number of correct anticipations by each $S$ for the Mediator Satiated pairs ($A_2$-$D_S$) from the corresponding total for the Mediator Nonsatiated pairs ($A_1$-$D_N$). If there is inhibition during learning of the Mediator Satiated pairs (i.e., $S$ has a smaller number of correct anticipations for the $A_2$-$D_S$ than the $A_1$-$D_N$ pairs), this difference will be positive. In view of Weitz's (1961) argument that conclusions based on verbal learning data often depend on the type of criterion measure used, it was decided to add four other measures in the present analysis and these are presented in Table 3. It can be seen that all six measures used in the comparison support the conclusion that semantic satiation of the connotative meaning of the mediator (C) increases the difficulty with which an A-D list is acquired after having established

A-B pairs. Furthermore, the proactive inhibition effect is noticeable not only as an increase in the number of trials required for acquisition, but also as an increase in both the number of intrusions and omissions made during acquisition. The same analysis for Group C is also given in Table 3. It will be remembered that Group C established A-X connections before learning A-D associations, and no satiation treatment was given. It can be seen that none of the six measures indicate differential difficulty of acquisition of the $A_1$-$D_N$ and $A_2$-$D_S$ pairs.

A closer examination of the data lends support to the main finding. If the $A_1$-$B_N$ pairs of List 1 had been originally better learned by experimental $S$s than the $A_2$-$B_S$ pairs, then the $A_1$-$D_N$ pairs of List 2 would have been easier to acquire than the $A_2$-$D_S$ pairs. Analysis of the total number of correct anticipations of the $A_1$-$B_N$ pairs minus the corresponding number for the $A_2$-$B_S$ pairs during learning of List 1 (to a criterion of one errorless repetition) yielded a mean of $-2.47$ ($SD = 15.51$; $t = 0.88$; $ns$). Thus, not only is there no reliable difference in the learning of $A_1$-$B_N$ and $A_2$-$B_S$ pairs in List 1, but also the $A_1$-$B_N$ pairs were somewhat more difficult to learn as indicated by the minus value of the mean difference. Consequently, the differential difficulty of A-D pairs under the satiation and nonsatiation treatments cannot be attributed to the initial differential difficulty of A-B pairs. Also, the fact that Group C did not exhibit superiority of the $A_1$-$D_N$ over the $A_2$-$D_S$ pairs of List 2 (Table 3) indicates that there was no intrinsic difference between the two sets of A-D pairs. Finally, $r = .42$ ($P < .02$) between the degree of semantic satiation shown by each experimental $S$ on the

satiated mediators ($C_S$) and the extent of inhibition shown in acquisition of $A_2$-$D_S$ pairs relative to the $A_1$-$D_N$ pairs. This is a most interesting finding since it shows that the extent of proactive inhibition for individual $S$s caused by satiation of the mediator is related to the degree of decreased meaningfulness of the mediator itself.

## Discussion

The finding that a significant facilitation effect is obtained in List 2 with Group E, but not with Group C, is a replication of the positive transfer effect of the mediation paradigm reported by several authors and will not be discussed further in detail. The specific contribution of this paper concerns the other finding reported, namely that significantly less facilitation is obtained when the mediator is satiated. Two possible mechanisms might be operating here: one is that the availability of the mediator is reduced, making the completion of the mediation sequence $B \rightarrow C \rightarrow D$ less probable; the other is that, given the assumption (see Staats, 1961) that some of the mediation reactions in the mediator C are also involved in B and D, the inhibition process might generalize to these two terms as well. As a result of either of these two processes, the subsequent acquisition of D will have been made more difficult. Furthermore, the amount of generalized inhibition or secondary extinction might be expected to be proportional to the degree to which the original stimulus word was inhibited. The significant positive correlation reported above is consistent with this expectation.

The authors view the present findings as further support for their interpretation of semantic satiation as a cognitive form of reactive inhibition having characteristics similar to extinction phenomena noted with conditioned responses (Jakobovits & Lambert, 1962a, 1962b; Lambert & Jakobovits, 1960). The present study also shows that the role of media-

tion in verbal learning can be studied from the point of view of generalization of inhibitory processes—an approach which is complementary to the positive semantic generalization paradigms used so far by previous investigators. Studies on the generalization of semantic inhibition, or, as in the present case, of mediated satiation, ought to prove useful in eliminating undesirable verbal habits, and perhaps even nonverbal habits, and may well provide a tool for the experimental manipulation of implicit verbalizations or meanings in studies of thinking and problem solving.

## SUMMARY

The present study was concerned with mediation in verbal transfer where generalized inhibition could be observed from one learning task to another. A proactive interference paradigm was arranged using the same method and material as Russell and Storms (1955) and McGehee and Schulz (1962). Assuming that mediation follows the sequence B → C → D, Ss first learned an A-B list, then the meaning of the inferred mediator, C, was reduced by a satiation procedure, and finally an A-D list was learned. It was shown that satiation of the mediator resulted in generalization of inhibition during learning of the A-D list. The results were discussed in the light of the existing theoretical models for mediated generalization.

## REFERENCES

ATHERTON, M. V., & WASHBURN, M. F. Mediate association studied by the method of inhibiting associations: An instance of the effect of "Aufgabe." *Amer. J. Psychol.*, 1912, **23**, 101–109.

HOWE, H. C. "Mediate" association. *Amer. J. Psychol.*, 1893, **6**, 239–241.

JAKOBOVITS, L. A., & LAMBERT, W. E. Semantic satiation in an addition task. *Canad. J. Psychol.*, 1962, **62**, 112–119. (a)

JAKOBOVITS, L. A., & LAMBERT, W. E. Semantic satiation among bilinguals. *J. exp. Psychol.*, 1962, **62**, 576–582. (b)

LAMBERT, W. E., & JAKOBOVITS, L. A. Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.*, 1960, **60**, 376–383.

McGEHEE, N. E., & SCHULZ, R. W. Mediation in paired-associate learning. *J. exp. psychol.*, 1962, **62**, 565–570.

PETERS, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18**, 20–48.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

STAATS, A. W. Verbal habit-families, concepts, and the operant conditioning of word classes. *Psychol. Rev.*, 1961, **68**, 190–204.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

WEITZ, J. Criteria for criteria. *Amer. Psychologist*, 1961, **16**, 228–231.

# WORK DECREMENT AND REMINISCENCE IN PIGEON OPERANT RESPONDING [1]

C. ALAN BONEAU AND SEYMOUR AXELROD

*Duke University*

The present study was prompted by the incidental observation that pigeons in a Skinner box seemed to exhibit some of the phenomena of human motor learning (see, e.g., Kimble, 1949). Specifically, after the pigeons had been trained to peck a key with a pattern projected on it, and not to peck when the key was blank (a relatively distributed work schedule), response rate declined over a series of successive positive periods (work decrement during relatively massed practice) and then increased markedly after a negative period in which response rate was low or zero (reminiscence). The experiment described here was a more systematic exploration of these phenomena; additionally, the question was asked whether the reminiscence effect is attributable merely to the rest during the negative period, or to inhibitory properties acquired by the negative stimulus during discrimination training.

## METHOD

The *S*s were 12 naive young-adult white Carneau pigeons, maintained at 75% ad lib. weight by manipulation of food intake. The *S*s were first key trained and then given 6 days of training on a 60-sec. variable interval (VI) schedule for 30 1-min. periods per day. Following this, *S*s were trained for 4 days, using a 60-sec. VI schedule for positive periods and no reinforcement for negative periods, to peck at a disk on which a pattern (circle with a gap at the top) was projected, and to withhold response when the disk was illuminated

but blank. Response periods lasted 60 sec., and alternated with 10-sec. off periods during which the box and key were unilluminated. Thirty periods per day were run, 15 positive and 15 negative, in random sequence with the restriction that no more than 2 positives or negatives were presented successively. For the next 4 (postdiscrimination test) days, 40 periods/day were run. During Periods 9, 18, 27, and 36 the key was either (D) dark (a rest period) or (L) lighted but blank (the negative stimulus) and reinforcement was withheld; during the remaining 36 periods, the positive stimulus was present and the 60-sec. VI schedule was maintained. The temporal positions of the dark and light periods were systematically assigned so as to distribute sequence effects.

## RESULTS AND DISCUSSION

In Fig. 1, average rates of responding for the 12 *S*s are plotted trial by trial for selected days of the training procedure. The first panel shows the relatively stable rate of responding on Day 6 (last) of VI training. The second and third panels show the rates to both the positive and negative stimuli for Day 1 and Day 4 (last) of discrimination training. As expected, there was a considerable warm-up effect at the beginning of each day.

The fourth panel of Fig. 1 shows the average rate per trial on the first day of the postdiscrimination training. Of interest are the high rates on the first trials of each block and the gradual decrease in rate until the interpolated blackout period or negative stimulus (Trials 9, 18, 27, and 36). The decrease is quite orderly. Individual *S*s showed a steady trial-by-trial decrease in rate, occasionally with no inversions. Least squares lines fitted to the last seven points of
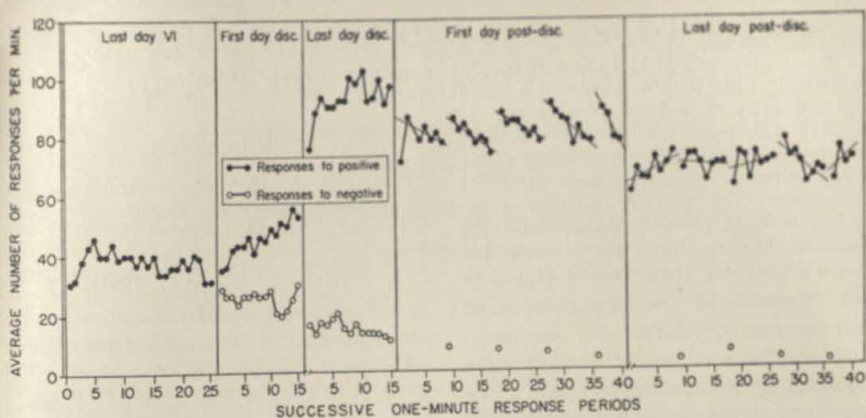
FIG. 1.  Rate of responding (pecks per minute) on selected days of the training procedure.

the first group of eight trials (to compensate for warm up), and to all points for the other blocks of trials, are also shown, and are the basis for the determination of reminiscence scores, defined as the difference between the extrapolated ninth-trial performance in one block and the fitted first-trial performance in the next.

Two differences in performance between the first day and the last day of postdiscrimination (fifth panel) should be noted.  First, the orderly reminiscence phenomenon occurring on the first day has disappeared by the fourth day.  Secondly, there is a change in level of response.  The average amount of reminiscence per bird on Day 1 of postdiscrimination, when tested against the null-hypothesized zero, yields a $t$ of 4.19, significant for 11 $df$ at the .01 level.  Response rate declined monotonically between the last day of discrimination and the last day of postdiscrimination ($F = 6.10$, $df = 4/44$, $P < .01$).

As outlined in the procedure, half of the interpolated 60-sec. rest periods in postdiscrimination were periods in which the negative stimulus was presented; the other half were periods in

which the key was not lighted.  The average difference in reminiscence over all 4 days following these two kinds of interpolated periods was .48 responses per min.  This difference results in a $t$ of .09, evidence consistent with the hypothesis that the reminiscence is due only to a low level of responding (rest) during the interpolated periods.

The experiment demonstrates the occurrence in the Skinner box of a number of the phenomena usually associated with human motor learning under conditions of massed and distributed practice.  The general superiority of distributed practice is evident during the latter stages of discrimination training when Ss are relatively inactive during the negative periods.  In the postdiscrimination periods occur the phenomena of work decrement (when the work periods are relatively massed) and reminiscence following a rest.  In addition, the average performance under the massed practice regime of the postdiscrimination period progressively decreases relative to distributed practice performance.  Warm-up decrement is not especially evident except at the beginning of each day's performance.  A final similarity to human motor performance is the decrease in the amount of reminiscence as massed practice training proceeds.

## Summary

Pigeons were trained in a Skinner box to discriminate a key with and without a figure projected upon it and exhibited rapid responding to the positive stimulus and little or no responding to the negative. No more than two positives or negatives were presented successively. The Ss were then given sequences of eight positives followed by one negative. The resulting performance exhibited phenomena characteristic of human motor learning performance under massed practice (warm-up decrement, temporary and permanent work decrement, and reminiscence).

## Reference

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.

# STUDIES OF DISTRIBUTED PRACTICE: XXII. SOME CONDITIONS WHICH ENHANCE RETENTION [1]

BENTON J. UNDERWOOD, GEOFFREY KEPPEL

*Northwestern University*

AND RUDOLPH W. SCHULZ

*University of Iowa*

The accumulated evidence indicates that in verbal learning some relatively high level of interference must be present before differences between massed practice (MP) and distributed practice (DP) will be observed. This law holds for differences in acquisition under MP and DP as well as for differences in retention following learning by MP and DP. However, the general statement relating interference and DP effects must be accompanied by subsidiary statements which relate particular DP phenomena to particular loci of the interference. A recent study (Underwood & Schulz, 1961) indicated that DP will facilitate acquisition only when interference occurs in acquiring or integrating response terms as such; DP will have no influence on retention when the interference is of this nature. On the other hand, when interference occurs in associating a response term to a stimulus term, DP may enhance retention but will have little effect on learning. It is this second subsidiary law with which the present experiments are concerned since these experiments explore further the conditions associated with better retention following DP than following MP.

In the study noted above, *S*s learned four paired-associate lists in which the stimulus terms were nonsense syllables and the response terms

were two-syllable adjectives. Because of the similarity across lists among the syllables, interference in associating the response terms with the stimulus terms increased as the number of lists increased. Each successive list required the learning of new responses to stimuli which were very similar to stimuli in previous lists. This similarity may be maximized by using the *same* stimuli in all lists in which case the four lists would be symbolized as A-B, A-C, A-D, and A-E. In line with the second subsidiary law, if DP is given on A-E, learning would not be facilitated but retention would be; DP appears to reduce proactive inhibition (PI) in retention when interference is of this type. Extinction and spontaneous recovery were postulated to be the basic mechanisms which reduced PI in the situation described above where interlist stimulus similarity was high. More particularly, it was assumed that in learning A-E the associations previously acquired (B, C, and D) to A were extinguished. Moreover, when A-E learning was by DP, it was assumed that the rest intervals allowed for some spontaneous recovery of the previously acquired associations; hence, subsequent learning trials would result in the re-extinction of these associations. Finally, it was assumed that these successive extinction-recovery cycles lead to a more permanent extinction of interfering tendencies under DP than under MP,

thus reducing PI in retention. To say that interfering tendencies are more permanently extinguished under DP than under MP is to say only that recovery will be slower and will never attain as high a level under DP as under MP.

The major purpose of the present experiments is to attempt to identify the associations which, according to this theory, are more permanently extinguished during DP than during MP. In an A-B, A-C paradigm it is possible to identify two different sets of associations which may be extinguished in learning A-C. The most obvious set of associations which might be extinguished are the A-B associations. As S learns C to A, the A-B association may become extinguished. These associations will hereafter be referred to as specific S-R associations. The response terms identify a second set of associations. The B response terms may be learned without also being associated with their particular verbal stimuli in the list. That is, response learning may occur independently of specific S-R learning. For this response learning the stimulus may be the general environmental situation or, as it will be called here, the stimulus may be a contextual one. In learning the A-C list it is possible that the associations between context stimuli and the B responses become extinguished and are replaced by the C responses. Thus, we may identify at least two kinds of associations which may be extinguished in learning A-C following the learning of A-B. The experimental question asked is whether the extinction of both types of associations is involved in the better retention following DP than following MP, or whether the effect can be attributed to one or the other class of associations. That extinction of both classes

of associations may occur seems to have been demonstrated by Barnes (1960). However, the relative permanence of the extinction of the two classes of associations and the relative recovery rates are unknown so that it is quite possible that the facilitating effect of DP on retention may be tied to the extinction of one class or the other rather than to the joint extinction of both classes.

In order to separate the effects of extinguishing the two classes of associations (specific S-R and contextual associations) a second paradigm is needed. This paradigm is known as the A-B re-paired paradigm in which on each successive list the same stimulus and response terms are used but the particular pairings differ from list to list. A study of this paradigm will show that in learning a second (or third, or fourth) list there are specific S-R associations which may be extinguished. However, since the same responses are used in each list, no extinction of contextual associations can occur, with the result that the contextual stimuli should remain associated for all lists. Thus, when the A-B, A-C paradigm is posed against the A-B re-paired paradigm it is seen that both have specific S-R associations which may be extinguished whereas only the A-B, A-C paradigm has contextual associations which may be extinguished. With these differences in the two paradigms in mind, the experimental situation to be used may be considered.

The Ss will learn four lists. In one case these lists will be A-B, A-C, A-D, A-E. In the other case there will be four A-B re-paired lists. To minimize the awkwardness in referring to these two paradigms the first will hereafter be called RD (responses different in each list) and the second RS (responses same in all lists). For both

paradigms DP is introduced in the learning of the fourth lists and the effect of DP on retention measured by having other groups in which the fourth list is learned by MP.

The results to be obtained in retention following MP and DP of List 4 for these two paradigms may demonstrate the need to limit application of the recovery-extinction theory to particular kinds of associations, i.e., either to specific S-R associations or to contextual associations. This is to say that the empirical results may be used to refine the gross theory. Three different possible empirical effects of DP on retention of List 4 for the two paradigms may now be stated along with the theoretical inferences to be drawn given each effect. (a) Effects of DP may be positive and equal for the RD and RS paradigms. The conclusion would be that extinction of specific S-R associations is responsible for the effect. (b) Effects may be positive for both paradigms but greater for RD than for RS. The conclusion would be that extinction of both specific S-R associations and contextual associations is responsible for the effect. (c) Effects may be positive for the RD paradigm with no effect for the RS paradigm. The conclusion would be that extinction of contextual associations is responsible for the effect.

Experiments 1 and 2 were directed specifically toward testing the above notions. Experiment 3, which deals only with the RD paradigm, was designed to investigate the influence of a different level of learning than used in Exp. 1 and 2 and to make some initial determinations of differences in length of the DP interval.

## METHOD

The three experiments have many procedural details in common. Therefore, Exp. 1 will be described in detail followed by a description of the changes introduced for Exp. 2 and 3.

### Experiment 1

*Design.*—Two parallel sets of conditions were used, one set being based on lists forming the RD paradigm and one set based on lists forming the RS paradigm. For each paradigm four lists were learned by each *S*. For the first three lists all *S*s within a paradigm were treated alike since the learning of these lists was imposed merely to build up interlist interference. For List 4, half the *S*s under each paradigm learned by DP (1-min. intertrial interval) and half by MP (4-sec. intertrial interval). Twenty-four hours after learning List 4, recall and relearning measures were obtained on it.

*Lists.*—Each list consisted of eight paired associates. The stimulus terms were nonsense syllables, the response terms two-syllable adjectives. For both paradigms only eight syllables are needed as stimuli for the four lists. The syllables used are those listed in Table 1 under List 1 of the previous experiment (Underwood & Schulz, 1961, p. 229). For the RS paradigm only eight different response terms are needed for the four lists. These terms were the adjectives given for List 1 of Table 1 in the previous study. To construct the four lists the stimulus and response terms were simply re-paired randomly with the restriction that a given pairing occur in only one list. For the RD paradigm, eight different adjectives were used for each list, these four sets being those in Table 1 of the aforementioned study. List 4 was exactly the same for *S*s under both paradigms, this list being the List 1 in Underwood and Schulz (1961) except, of course, in the present study the adjectives were response terms and the syllables were stimulus terms. The sequence of Lists 1–3 was the same for all *S*s within a paradigm, i.e., they were *not* counterbalanced.

*Procedure.*—Lists were presented at a 2:2-sec. rate with anticipation learning throughout. Each list for *S*s in the RD paradigm was presented for 12 anticipation trials with *S* instructed to give as many correct anticipations as possible on each trial. Four different orders of items were used for each list with each order being used as a start order about equally often. Lists 1 and 2 were presented on Day 1; Lists 3 and 4 on Day 2, and 5 relearning trials were given on List 4 on Day 3, 24 hr. after Day 2. Relearning was by MP
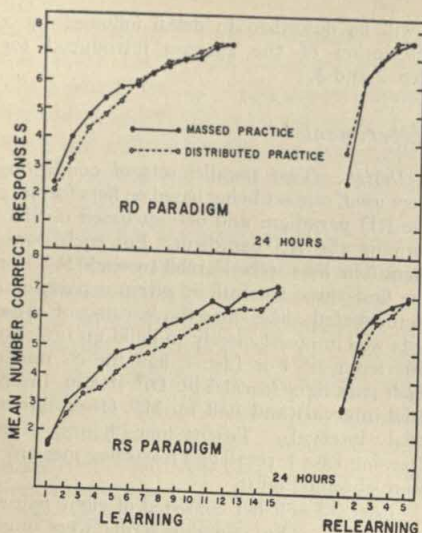
FIG. 1. Learning and relearning curves for Exp. 1.

for all groups. Approximately 30 sec. elapsed between the learning of the two lists on a given day. The DP intervals on List 4 were introduced after each trial and were filled with symbol cancellation. The procedures for Ss in the RS paradigm were exactly the same as for those outlined above except that 15 anticipation trials were given on each list. Since the RS paradigm produces more interference than the RD paradigm, 15 trials were used for the RS paradigm (as compared with 12 for the RD) in order to make the level of learning under the two paradigms roughly comparable.

The Ss were all college students, and, while all were not naive to verbal-learning experiments, none had previously learned syllable-adjective pairs. For the RD paradigm 32 Ss learned List 4 by DP and 32 by MP. For the RS paradigm there were 30 Ss in each group.

## Experiment 2

As in Exp. 1, both paradigms were studied. The changes from Exp. 1 were as follows: (a) Lists 1–3 were given on Day 1, List 4 was given on Day 2, and List 4 was relearned for 5 trials on Day 3, 24 hr. after Day 2. (b) Under both paradigms 12 anticipation trials were given on all lists. (c) MP was 4 sec. between trials while DP consisted of a 3-min. interval after Trials 1, 3, 5, 7, 9, and 11. (d) All four groups (MP and DP groups for each paradigm) had 32 Ss.

## Experiment 3

In Exp. 3 only the RD paradigm was studied. All four lists were presented during a single session but each was presented for only 5 anticipation trials. Three groups of 25 Ss each were used, the groups being differentiated only by the intertrial interval on List 4. One group had the usual MP (4-sec. intertrial interval), a second had 60 sec. between trials, and a third had 180 sec. between trials. Again List 4 was given 5 relearning trials 24 hr. after original learning.

## RESULTS

*Experiment 1.*—The equivalence of the learning ability of the two groups of Ss within each paradigm can be gauged by the learning scores for Lists 1–3 since the conditions were identical for all Ss for these lists. For the RD paradigm, the mean total correct anticipations for Lists 1–3 in order for the MP Ss were 64.09, 64.00, and 72.31. For the DP Ss the corresponding values were 63.38, 63.53, and 73.03. For the RS paradigm the values for the MP Ss were 79.57, 80.67, and 76.73 and for the DP Ss, 78.40, 80.53, and 75.23.

The acquisition and relearning curves for Ss in both paradigms for List 4 are shown in Fig. 1. For both paradigms learning is somewhat slower under DP than under MP; this is true throughout the 15 trials for RS lists and for the initial part of learning for the RD lists. This finding confirms

that of an earlier study (Underwood & Schulz, 1961) and also conforms to the extinction theory in that the DP interval should allow interfering associations to recover, thus impeding acquisition.

Looking next at the recall trial it can be seen that for the RS paradigm recall is almost identical for both MP and DP. The slight difference in the relearning curves in favor of MP probably reflects the small difference present in original learning. It must be concluded that DP does not facilitate retention of lists forming the RS paradigm. For the RD paradigm, however, recall is better under DP than under MP. The exact values are 2.59 for MP and 3.66 for DP ($t = 2.28$, $.05 > P > .02$). If loss scores are used as described in the previous study (Underwood & Schulz, 1961, p. 231–232), in order to adjust for any differences in original learning, the evaluation does not change ($t = 2.32$). It is concluded that DP in the RD paradigm facilitates recall after 24 hr. The effect, however, is very transitory since no difference in the subsequent four relearning trials is evident.

Certain expectations concerning overt errors follow from the extinction hypothesis. First, more intrusions from previous lists should occur in learning List 4 under DP than under MP. In the RS paradigm, the definition of an intrusion is somewhat ambiguous since the same responses occur in all lists. However, the nearest approximation would be to consider only responses given to a stimulus with which that response had been paired in previous lists. Approximately two-thirds of all errors made in learning List 4 were of this nature. And, while more of these (a total of 387) occurred under DP than under MP (347), statistically speaking, the difference is far from significant. For

the RD paradigm, 13 responses from List 1–3 occurred in learning List 4 under DP, and 7 under MP. While this difference is in accordance with the theory, the numbers are so small that little should be made of the effect.

A second expectation from the theory is that fewer intrusions should occur in relearning following DP than following MP. For the RS paradigm 132 intrusions (as per the definition given above) occurred in relearning following MP and 138 following DP. Clearly, there is no evidence in this paradigm of a more permanent extinction occurring under DP than under MP. For the RD paradigm, 44 intrusions occurred during relearning following MP and 24 following DP. These data are in conformance with the theory.

*Experiment 2.*—In this experiment, Lists 1–3 were learned on Day 1 with List 4 being learned on Day 2, and the DP interval for List 4 was 3 min. after every other trial.

In learning Lists 1–3 the mean total correct responses for the MP *S*s in the RD paradigm were 74.59, 69.53, and 74.13. For the DP *S*s the corresponding values were 69.78, 64.34, and 73.13. The MP *S*s learning the RS lists showed means of 68.38, 60.97, and 67.00, while the comparable values for the DP *S*s were 69.13, 56.56, and 64.25. Thus, the two groups within each paradigm are fairly comparable in learning ability.

The performances in learning and relearning List 4 are plotted in Fig. 2. Again the *S*s learning RD List 4 under DP show inferior performance to those learning under MP. For the RS paradigm, the picture is a little different. After nearly every rest interval the DP *S*s show inferior performance to the MP *S*s but on the immediately succeeding trial the per-
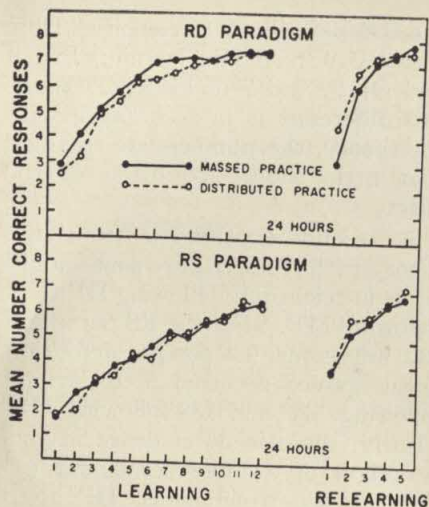
FIG. 2. Learning and relearning curves for Exp. 2.

formance of the two groups is essentially equivalent. Thus, while the rest interval appears to inhibit performance, the overall learning rate does not appear to be seriously impaired.

The recall of the DP $Ss$ under the RD paradigm is again superior to the MP $Ss$. The mean values are 3.13 for MP $Ss$ and 4.53 for the DP $Ss$ ($t = 3.11$, $P < .01$). Clearly, DP facilitates recall of the fourth list for this paradigm. The DP $Ss$ are also superior to the MP $Ss$ for the first four trials of relearning. However, the difference across the five trials in terms of mean total correct responses (1.85) is not significant ($t = 1.37$).

For $Ss$ learning List 4 of the RS paradigm no difference of consequence is apparent in recall and relearning. While recall favors the DP $Ss$, the difference of .11 items gives a $t$ of only .22. Thus, as in Exp. 1, it is seen that recall of List 4 is facilitated in the RD paradigm and no appreciable effect is noted for the RS paradigm.

No difference of consequence in number of intrusions in learning under

MP and DP for the RS paradigm was noted; the same was true for relearning. For the RD paradigm 13 intrusions were recorded for the MP $Ss$ learning List 4 and 9 for the DP $Ss$. In this case the difference is in the opposite direction from the difference found in Exp. 1 and is not in accordance with extinction theory. For relearning, however, MP $Ss$ made 55 intrusions and DP $Ss$, 28, which, as in Exp. 1, is in line with expectations from the theory.

*Experiment 3.*—In this experiment only the RD paradigm was used. Lists 1–4 were presented for five acquisition trials each, with three subgroups of 25 $Ss$ each having 4, 60, or 180 sec. as the intertrial interval on List 4. The mean total correct responses given on Lists 1–3 by the 4-sec. $Ss$ were 17.76, 14.92, and 19.52, respectively. The comparable scores for the 60-sec. $Ss$ were 17.80, 14.92, and 20.28, and for the 180-sec. $Ss$, 17.88, 14.96, and 21.32.

The learning and relearning curves for List 4 for the three groups are shown in Fig. 3. Again it is to be noted that during learning the performance of the two DP groups is inferior to that of the MP group. Both DP curves are consistently lower than the MP curve throughout the five trials. Again, however, the dif-
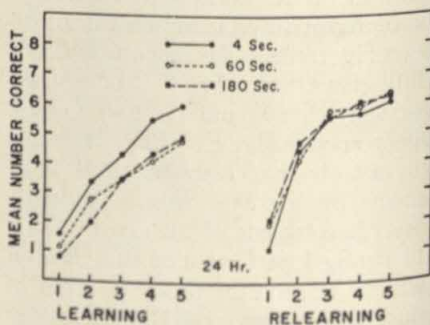


FIG. 3. Learning and relearning curves for Exp. 3.

ferences are just short of significance at $P = .05$ ($F = 2.91$, $df = 2/72$). Nevertheless, in view of the fact that for this paradigm DP has been consistently inferior to MP in learning in all three experiments, the inhibitory effect may be taken to be reliable.

The mean number of responses correctly recalled after 24 hr. were 1.20, 2.04, and 2.12 for 4, 60, and 180 sec., respectively ($F = 1.95$, $df = 2/72$, $P > .05$). However, it seems clear that adjustments must be made to compensate for differences in level of learning originally attained by the three groups. When this adjustment is made by the method noted in the earlier paper (Underwood & Schulz, 1961), the $F$ becomes 21.40 ($df = 2/72$, $P < .001$). Therefore, it may again be concluded that for the RD paradigm distributed practice reduces PI in retention.

In the results of Exp. 1 and 2 it was noted that the facilitation in retention for the RD paradigm produced by DP was largely limited to the first recall trial. In the present data the effect is much less transitory; in spite of the fact that MP Ss were appreciably better in learning, all groups are about equal in relearning. This effect of DP on relearning may be estimated by subtracting for each S the number of correct responses given on the five learning trials from the number of correct responses given on the five relearning trials. The means for these differences were 2.16, 7.84, and 9.20 for the 4-, 60-, and 180-sec. Ss, respectively. The $F$ (12.60) is beyond the .01 level. The potency of the PI effects on the MP Ss can be further described by noting that of the 25 Ss, 11 showed performance in relearning which was inferior to that shown during learning. By the same token the reduction in PI effects produced by DP can be seen by noting that

only one of the 60-sec. Ss performed more poorly in relearning than in learning and none of the 180-sec. Ss performed so.

In learning List 4 the 4-sec. Ss gave 7 intrusions, i.e., responses which were appropriate for an earlier list. The 60-sec. Ss gave 32 such responses and the 180-sec. Ss gave 11. In the 4-sec. condition all 7 intrusions were given by a single S, whereas 16 Ss gave intrusions under the 60-sec. condition and 4 under the 180-sec. condition. Thus, while more intrusions were made under the DP conditions than under the MP condition, the fact that the 60-sec. Ss gave more intrusions than did the 180-sec. Ss would not be anticipated by the extinction hypothesis.

The intrusions during relearning totaled 34, 33, and 25, for the 4-, 60-, and 180-sec. Ss, respectively. The theory predicts more intrusions for MP Ss than for DP Ss. While the values indicate this to be the case the differences are so small that they do not allow any strong support for this aspect of the theory.

## DISCUSSION

The results of Exp. 1 and 2 will be evaluated first. The basic findings were as follows. When Ss learn four lists with the RS paradigm, DP on List 4 does not facilitate the 24-hr. retention of this list. But, when Ss learn four lists with the RD paradigm, the retention of List 4 is facilitated by learning the list under DP conditions. In terms of the argument advanced in the introduction, the implication of these facts is that retention of a list learned by DP is facilitated only when the paradigm allows extinction of contextual associations (associations between the general environment and the response terms). Both paradigms studied allow, presumably, extinction of specific S-R associations but only the RD paradigm allows for extinction

of contextual associations. Thus, the notion that DP, by allowing for successive recovery-extinction cycles, leads to a more permanent extinction of interfering tendencies appears to be supported for contextual associations only.

The question may be raised as to why the successive-extinction hypothesis is not supported in the case of specific S-R associations in view of the fact that the extinction of such associations seems to have been demonstrated (Barnes, 1960; Barnes & Underwood, 1959). One possibility is that recovery of such associations occurs very slowly; thus, the short DP intervals used here may not in fact allow for recovery-extinction cycles. And in fact there is evidence (e.g., Briggs, 1954) that recovery of such associations is indeed very slow. A barrier to the acceptance of this position is the fact that in learning List 4 in the RS paradigm there was evidence for *some* process which was impeding performance under DP. It is reasonable to think that this could represent the recovery of error tendencies of some sort. Yet, it is possible that the interference which appears to increase with DP of List 4 in both paradigms is not representative nor symptomatic of the process which produces the better retention following DP of List 4 under the RD paradigm. The inconsistency from experiment to experiment of the differences in intrusion frequency between MP and DP on List 4 might argue for such a position. Therefore, although slower learning of List 4 by DP than by MP is consistent with an extinction hypothesis, the slower learning need not necessarily be taken to mean that better retention occurs as a consequence.

Empirically speaking, there is a relatively simple principle which may be stated which will summarize the situations in which DP may be expected to facilitate retention. As noted in the introduction, to expect any facilitation in retention by DP during learning requires first that appreciable interlist interference be present. Given this situation, the principle is that whenever the response terms of the previously learned associations producing the interference are *not* present in the list being learned, DP will facilitate the retention of the list. This empirical induction not only conforms to the paradigms and findings of the present experiments but also will handle other findings. For example, when there is high intralist similarity among syllables within a list, no effect on retention is noted if the list is learned by DP (Underwood & Richardson, 1958). The interference falling on a given association in this situation is caused by letter sequences which are appropriate or correct for other associations in the list; that is, the response units of the associations producing the interference are correct for other items in the list. The same situation holds for the RS paradigm in the present studies. To account for this empirical generalization—the generalization that DP will facilitate retention only when the response terms for the associations producing the interference are not in the list being learned—we have used the notion of recovery-extinction cycles leading to a more permanent extinction of contextual associations.

Why these contextual associations "behave" differently than other associations (e.g., specific S-R associations) is not known. Of course, it should be remembered that the present experiments have only scratched the surface in terms of studying what appear to be the relevant variables and their interaction. Degree of learning of the interfering lists, degree of learning of the list being interfered with and to be recalled, length of DP interval, and length of retention interval, all should be pertinent variables. Clearly they are pertinent in terms of the extinction-recovery theory but they undoubtedly would also be judged to be of importance by any careful empirical analysis.

The final point of discussion concerns the results of Exp. 3. In this experiment the RD paradigm was used and, in conformance with the results of Exp. 1 and 2 with this paradigm, facilitation in retention of List 4 was observed following learning by DP. The effects on retention were rather substantial in that

relearning was enhanced following DP. Clearly, PI was reduced by DP. It had been expected that the 3-min. DP interval would produce more facilitation than would the 1-min. DP interval. There was no strong evidence of differences in the effect of these two conditions. It is true that relearning was a little better following a 3-min. DP interval than following the 1-min. interval. Also, analyses of particular items showed that items having the greatest number of correct anticipations at the end of learning of List 4 produced higher recall following 3-min. DP than following 1-min. DP. Still, the effects were not as great as anticipated. In terms of the extinction-recovery notion, it might be argued that 1 min. was sufficient for the recovery of all interfering tendencies for the relatively low degree of learning of the interfering lists used. If this is true, a more direct relationship between retention and length of intertrial interval would be expected if the degree of learning of the interfering associations were higher. A comparison of the results for the RD paradigm of Exp. 1 and 2 would tend to support this notion. In these experiments the degree of learning of the interfering associations was much stronger than in Exp. 3. The 3-min. interval of Exp. 2 gave somewhat greater facilitation in retention than did the 60-sec. DP interval used in Exp. 1. However, this is by no means a clear test since other factors also varied between the two experiments. In any event, some conviction is held for the notion that in most situations a direct relationship between length of intertrial interval and facilitation in retention will be found.

## Summary

Three experiments were performed to study retention following massed (MP) and distributed practice (DP) when interlist interference was high. In Exp. 1 and 2 half the Ss learned four lists of paired associates (syllable-adjective pairs) in which the stimuli were identical across all lists but with the responses different in each list (RD paradigm). The other Ss learned four lists in which the stimuli and responses were identical across all lists but with different pairings for each list (RS paradigm). Distributed practice was introduced in learning List 4 with retention of this list measured after 24 hr. In Exp. 1 the DP interval was 60 sec. between each trial while in Exp. 2 the interval was 3 min. after every other trial; MP consisted of 4 sec. between trials for both experiments.

The results were essentially the same for Exp. 1 and 2; DP facilitated recall of List 4 only for the RD paradigm. These findings indicate that DP will facilitate retention only when the response terms of previously acquired associations producing the interference are not present in the list being learned by DP. Theoretically, the results imply that DP allows for a more permanent extinction of contextual associations but does not influence specific S-R associations.

Experiment 3 used only the RD paradigm with a low degree of learning of all four lists and with intertrial intervals of 4, 60, and 180 sec. on List 4. Distributed practice markedly facilitated both recall and relearning of List 4. However, no appreciable difference was noted in the results for the 60- and 180-sec. intervals.

## References

Barnes, J. M. "Fate" revisited. Unpublished doctoral dissertation, Northwestern University, 1960.

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

Underwood, B. J., & Richardson, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56**, 213–219.

Underwood, B. J., & Schulz, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235.

# THE ROLE OF RESPONSE SIMILARITY IN PROACTIVE INHIBITION [1]

## KENT M. DALLETT [2]

*University of California*

It has generally been supposed that interlist response similarity is an important determinant of interlist interference. However, while variations in response similarity have several times been shown to affect retroactive inhibition (RI) (Slamecka & Ceraso, 1960), effects of interlist similarity on proactive inhibition in retention (PI) have not yet been demonstrated. Young (1955) failed to find significant differences in PI as a function of response similarity, although he did find the expected differences in RI. Similar findings with respect to PI were reported by Morgan and Underwood (1950), the only significant PI in their study resulting from the use of identical stimuli and dissimilar responses. In each of the studies mentioned, however, there was an insignificant trend suggesting an effect similar to that generally found in RI; i.e., recall improved as response similarity increased. This trend was directly opposed to the differences predicted by Young (1955), who supposed that as response similarity increased, greater generalized strength would be added to List 1, increasing its potential for interfering with List 2 recall.

Recently, Barnes and Underwood (1959) have proposed separate mechanisms for the learning of similar and dissimilar responses in List 2. When the responses in List 2 are dissimilar to those on List 1 (with identical stimuli: A-B, A-C), List 1 responses are unlearned during List 2 learning; while with similar responses, it is suggested that S makes use of the List 1 response as a mediator. Thus, in learning A-B' after A-B, S is assumed to be learning A-(B)-B', although Ss report the dropping-out of the mediator as learning progresses. Accepting Barnes and Underwood's analysis, and assuming that PI results from the response competition produced when unlearned List 1 responses spontaneously recover in strength, one might be led to expect considerable PI in the A-B, A-C paradigm, but little or no PI in the case of A-B, A-B' (Postman, 1961). This analysis can readily be extended to intermediate degrees of response similarity by assuming that as response similarity increases from dissimilarity, unlearning gradually gives way to mediation. In this way, the data of Morgan and Underwood (1950) and of Young (1955) can be explained.

However, we do not know that the unlearning-and-recovery sequence is the only mechanism involved in PI. It is possible, for example, that similar responses may lead to loss of differentiation (Underwood, 1945) with the passage of time. Thus, Ss may recall both responses, but may not remember which response belongs in List 2. One might, on the other

hand, expect proactive facilitation if the List 1 mediator protects List 2 responses from the extraexperimental interference to which a control group may be subject. These possibilities hold for delayed recall only, little PI being expected at short intervals in any case. Indeed, since PI appears to increase with time, it would seem advisable to investigate the effects of response similarity on PI using retention intervals considerably longer than the 20-min. retention intervals employed by both Morgan and Underwood (1950), and Young (1955).

In the experiment to be reported, independent groups of $S$s learned two lists with identical stimuli ($S_I$) and highly similar ($S_I R_1$), less similar ($S_I R_2$) or dissimilar ($S_I R_N$) responses, different $S$s being tested for recall of List 2 either 30 sec. or 48 hr. after learning. In addition, 30-sec. and 48-hr. groups were tested under each of two control conditions: (a) the standard PI control group, having no first list (No PL), and (b) a group having dissimilar stimuli and responses on the two lists ($S_N R_N$). This second control condition served as a control for warm-up and learning-to-learn effects in the learning of List 2, and also was intended to sample the low-similarity end of the continuum of stimulus similarity.

## METHOD

*Materials and apparatus.*—The words used as responses were chosen from Haagen's (1949) norms. Subjects in all conditions learned the same List 2. Corresponding to a given word on List 2, words were chosen for three different first lists so as to be highly similar ($R_1$), less highly similar ($R_2$), or dissimilar ($R_N$) to the List 2 word. The $R_1$ and $R_2$ lists had mean Haagen similarities of 1.40 and 3.36, respectively. The $R_N$ list was made up of Haagen words not in the same category as the corresponding List 2 response. In addition, an attempt was made to match corresponding words for familiarity using

RESPONSES USED IN THE EXPERIMENT

| List 2 | List 1 | | |
| | $R_1$ | $R_2$ | $R_N$ |
| --- | --- | --- | --- |
| AGILE | NIMBLE | ALERT | UNKIND |
| BELOVED | CHERISHED | VALUED | PETTY |
| COMPLETE | ENTIRE | PERFECT | HEAVY |
| CRAFTY | CUNNING | STEALTHY | PRIOR |
| DISTANT | REMOTE | FURTHER | SPOKEN |
| DECEASED | LIFELESS | EXTINCT | OBSCENE |
| FRUITFUL | FERTILE | PREGNANT | IMPURE |
| HAUGHTY | SNOBBISH | SCORNFUL | SHAKY |
| LIQUID | FLUID | JUICY | FOREMOST |
| SACRED | HOLY | TABOO | IDLE |
| SHINING | GLEAMING | SPARKLING | CONCEALED |
| WICKED | EVIL | VICIOUS | DAINTY |

Haagen's norms, and for frequency of usage using the Lorge magazine count (Thorndike & Lorge, 1944). Thus, the three sets of List 1 responses, as well as the List 2, or "standard" set, were matched both in mean frequency and familiarity and in frequency and familiarity of corresponding words. The responses are presented in Table 1.

The stimuli were CVC trigrams of 93–100% Glaze association values (Hilgard, 1951). Two sets of 12 trigrams were selected so as to minimize both interlist and intralist similarity. Only in two instances did syllables on the two lists have two letters in common, and within each list there were 16 repetitions of letters.

Stimuli and responses were combined in four different pairings, making a total of 20 lists. Each list was presented in four different orders. The lists were presented on a Phipps and Bird memory drum at a 2:3 sec. rate, with 6 sec. between trials.

*Procedure.*—Subjects were seated before the memory drum and read standard paired-associate learning instructions. On the first presentation of each list they were not required to respond, but on subsequent presentations they were encouraged to try to anticipate as many of the words as possible, and it was made clear that there was no penalty for guessing. List 1 was then presented nine times. The list was changed, and $S$ was told that he was to learn List 2 "in the same way." The change of lists required 30–40 sec. Nine trials were given on List 2. Following List 2, 30-sec. $S$s were encouraged to stretch their legs for a few seconds: after approximately 15 sec. instructions for recall were given them. Those $S$s assigned to the 48-hr. recall groups were told to return in 2 days for the "second hour of the experiment." The recall instructions differed slightly in

wording for the 30-sec. and 48-hr. groups. For each condition, however, it was emphasized that they were to recall the second list, and that they should begin with the very first syllable they saw. The differences in wording were minor, consisting mainly of greater emphasis on the fact that List 2 was in question for the 48-hr. Ss, and presentation of the task as involving possible effects of a "brief pause" for the 30-sec. Ss.

*Subjects.*—The Ss were students from the introductory course in psychology at the University of California. They participated in order to fulfill a course requirement, and were naive with respect to verbal learning. They were assigned to conditions at random prior to their appearance, except that 30-sec. and 48-hr. Ss were tested in alternation.

Since the experiment was concerned with the effects of List 1, it was felt advisable to discard Ss who did not meet a minimal criterion of List 1 learning within the nine trials allowed. On the basis of the pilot Ss, it was decided to discard Ss who had not achieved at least five correct anticipations on any trial. Seven Ss were dropped for not reaching this criterion. An additional 9 Ss were dropped for failure to follow instructions; 6 of these admitted not trying to anticipate overtly on early trials, 2 were attempting to learn serial order, and 1 did not anticipate items which had previously been correctly anticipated. While these failures to follow instructions may simply result from the Ss in question being slow learners, they could
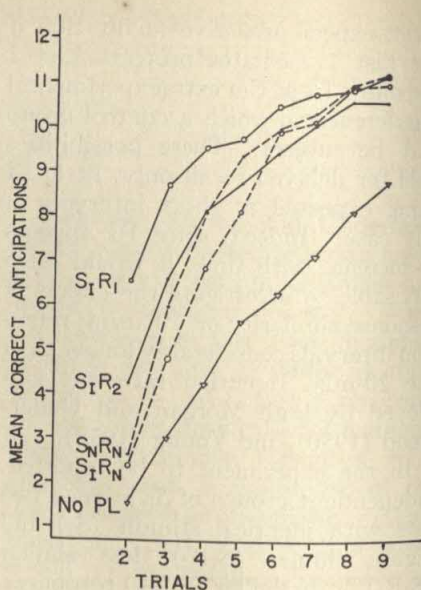


FIG. 1. List 2 learning curves.

usually do quite well when the instructions were explained to them. An additional 3 Ss were dropped for E error or apparatus failure. The number of Ss discarded for any one of the above reasons in any one similarity condition (pooling 30-sec. and 48-hr. groups) did not exceed 3. A total of 120 Ss, 12 in each of the 10 groups, was retained.

## RESULTS

*List 1 learning.*—An analysis of variance of total correct anticipations for the eight anticipation trials of List 1 revealed no differences significant at the .05 level. A similar analysis of intralist errors also revealed no differences. The mean of total correct anticipations for all groups combined was 49.95, with SDs ranging from 11.37 to 18.21. For intralist errors, the mean was 8.22, with SDs from 2.47 to 7.44. The groups with high response similarity ($S_IR_I$) made a total of seven interlist errors (words scheduled to appear in List 2) before ever having seen List 2.

*List 2 learning.*—An analysis of variance of total correct anticipations

TABLE 2

CORRECT RESPONSES AND ERRORS IN LIST 2 LEARNING

| Condition | | Correct Anticipations | | Intra-list Errors % | Inter-list Errors % |
|---|---|---|---|---|---|
| | | Mean | SD | | |
| No PL | 30 sec. | 47.08 | 15.71 | 21.1 | — |
| | 48 hr. | 41.25 | 13.24 | 16.8 | — |
| $S_NR_N$ | 30 sec. | 69.08 | 14.80 | 14.5 | 0.6 |
| | 48 hr. | 68.00 | 8.74 | 15.6 | 0.0 |
| $S_IR_N$ | 30 sec. | 65.42 | 12.00 | 7.5 | 1.4 |
| | 48 hr. | 62.92 | 12.51 | 20.3 | 0.7 |
| $S_IR_2$ | 30 sec. | 63.17 | 15.75 | 16.1 | 5.1 |
| | 48 hr. | 73.75 | 12.13 | 14.3 | 4.6 |
| $S_IR_I$ | 30 sec. | 77.08 | 8.28 | 16.9 | 18.0 |
| | 48 hr. | 78.67 | 9.82 | 21.0 | 15.6 |

similar to that employed for List 1 learning was carried out. Groups scheduled for 30-sec. and 48-hr. recall, not yet differentially treated, did not differ significantly from one another. There were significant differences among the groups as a function of interlist similarity ($F = 23.738$, $df = 4/110$, $P < .005$), and these differences, while small, were in the expected direction (Table 2). All groups showed net positive transfer with respect to the group with no PL. The $S_I R_N$ groups, while showing positive transfer with respect to No PL, were slightly inferior to the $S_N R_N$ groups. This is in agreement with several recent studies which have shown that $S_I R_N$ may produce negative transfer only with respect to a group with equivalent warm-up and learning-to-learn experience (e.g., Besch & Reynolds, 1958; Spiker & Holton, 1958). Examination of the learning curves (Fig. 1) indicates that after the nine trials, all groups with prior learning are within one correct anticipation of one another, while the No PL $S$s reached a criterion comparable to that reached by the other groups on List 1.

Since differences were found among groups in correct anticipations, all error measures were corrected for opportunity. The total number of intralist errors for each $S$ was expressed as a proportion of that $S$'s opportunities for error: each stimulus presentation not resulting in a correct anticipation was considered one such opportunity. The resulting proportions were submitted to an arc-sine transformation, and an analysis of variance was carried out. No significant differences in intralist errors were found. In the case of interlist errors, which were extremely infrequent in all groups except those with similar responses ($S_I R_1$, $S_I R_2$), the total number of such errors was expressed as a proportion of total opportunities for each group. These data are presented as percentages in Table 2: the percentages, when plotted against decreasing interlist similarity, yield decreasing regular functions, with the 30-sec. groups closely paralleling the 48-hr. groups. A similar gradient results from plotting the number of $S$s making at least one interlist error. In the $R_1$ groups, 21/24 $S$s made at least one interlist error, while in the $R_2$ groups 12/24 $S$s made such errors, and in the $S_I R_N$ groups only 5/24 $S$s made interlist errors.

*Recall.*—In the evaluation of List 2 recall, two methods of estimating differences attributable to differences in List 2 learning have been used. In one, the 30-sec. recall test is used to estimate the strength of List 2 at the end of learning: any difference between 30-sec. and 48-hr. recall scores is attributable to the differential time of testing. A similar estimate results from the other method used, the successive probability analysis (Underwood, 1954, 1956 unpublished[3]). These methods do not allow one to say whether groups which have attained different criteria forget different amounts because of the experimental treatment, or simply because differential amounts of loss are characteristic of different criteria. The ambiguity resulting from different groups having different performance levels at the end of learning arises primarily with respect to comparisons of the group having no PL with the other groups, the No PL $S$s having ended List 2 learning at a lower level of performance. However, there is no pressing reason to suppose that $S$s

[3] Unpublished manuscript by B. J. Underwood entitled, "Strength of association and forgetting."

## TABLE 3
### Correct Responses and Errors in Recall

| Condition | | Correct Anticipations | | Intra-list Errors % | Inter-list Errors % |
|-----------|---|------|------|------|------|
| | | Mean | SD | | |
| No PL | 30 sec. | 8.66 | 2.53 | 16.2 | — |
| | 48 hr. | 6.17 | 1.53 | 9.3 | — |
| $S_N R_N$ | 30 sec. | 11.00 | 1.71 | 8.3 | 0.0 |
| | 48 hr. | 7.42 | 2.27 | 6.4 | 0.0 |
| $S_I R_N$ | 30 sec. | 11.17 | 1.19 | 10.0 | 0.0 |
| | 48 hr. | 5.33 | 3.05 | 2.5 | 10.0 |
| $S_I R_2$ | 30 sec. | 9.83 | 2.04 | 13.5 | 9.6 |
| | 48 hr. | 7.50 | 2.07 | 5.5 | 11.1 |
| $S_I R_1$ | 30 sec. | 10.92 | 1.00 | 19.2 | 19.2 |
| | 48 hr. | 7.25 | 2.09 | 3.5 | 18.4 |

who have reached a lower criterion would ordinarily forget *less* than Ss who have reached a higher criterion, which appears to have been the case in these data.

Turning first to the comparison of 30-sec. and 48-hr. recall, an analysis of variance revealed that the interaction of retention interval and conditions was significant at the .05 level ($F = 2.822$, $df = 4/110$). The main effects of Conditions ($F = 3.053$, $df = 4/110$) and Retention Interval ($F = 93.04$, $df = 1/110$) were also significant, indicating merely that different conditions performed differently in retention (as in learning), and that forgetting occurred. The mean recall scores are presented in Table 3. They suggest that a good part of the significant interaction may be due to the fact that the loss in $S_I R_N$ is larger than the loss in any other group. When the interaction is partitioned, $S_I R_N$ accounts for most of the variance, the difference among the other groups in amount lost being insignificant.

For the successive probability anal-

ysis, the probability of a correct response on Trial 9 following 1, 2, . . . 7 correct anticipations on Trials 2-8 was obtained for each condition, using the combined learning data of the 30-sec. and 48-hr. groups. Since the probability of a correct anticipation following seven correct anticipations was .95 or better in all groups, the probability of a correct response following eight correct anticipations was assigned an arbitrary value of 1.00.

Loss scores were obtained for each 48-hr. S individually, by computing an expected recall score on the basis of the probability analysis, and subtracting from this value the obtained recall. These loss scores (Table 4) were subjected to an analysis of variance which confirmed the results of the first analysis of raw recall scores. The conditions differ in amount lost ($F = 4.53$, $df = 4/55$, $P < .01$), and the differential loss appears to result mainly from the greater loss in Cond. $S_I R_N$, with no difference among the other groups. The logic of the experiment, however, justifies one selected comparison between the No PL control group and the other groups (with $S_I R_N$ excluded). This difference is significant at the .05 level ($F = 4.868$, $df = 1/55$), but since it is a selected comparison, one would probably insist upon significance at the .01 level or better. Finally, a test described by Snedecor (1957, p. 251) which allows for the

## TABLE 4
### Means and SDs of Expected Recall, Based on Probability Analysis, Minus Obtained Recall

| Measure | Condition | | | | |
|---------|-----------|------|------|------|------|
| | No PL | $S_N R_N$ | $S_I R_N$ | $S_I R_1$ | $S_I R_1$ |
| Mean | 2.61 | 4.06 | 6.05 | 3.88 | 4.32 |
| SD | 1.31 | 1.97 | 2.98 | 1.39 | 1.97 |

change in probabilities attendant upon repeated testing of selected groups also indicated that the only significant difference in the loss scores is between $S_I R_N$ and the other groups.

Thus, only in Cond. $S_I R_N$ has PI been demonstrated. This finding is the same as that reported by Morgan and Underwood (1950) with 20-min. recall.

*Intralist and interlist errors.*—Since errors were not considered to be frequent enough to justify correcting each $S$'s errors, total errors of each kind for each group of $S$s were expressed as a proportion of total opportunities for that group. As in learning, there does not seem to be any obvious systematic relationship between intralist errors and recall.

Interlist errors as a percentage of opportunity *increase* from 30 sec. to 48 hr. for $S_I R_N$ (Table 3) while the two groups with similar responses manifest the same percentage of interlist errors in the two recall tests. The $S_N R_N$ groups made no interlist errors in recall. Since 48-hr. recall scores were lower than 30-sec. recall scores, $S_I R_N$, $S_I R_2$, and $S_I R_1$ all showed an increase in the absolute frequencies of interlist errors, but only in $S_I R_N$ was this increase out of proportion to the increase in opportunities for error.

*Transfer and recall as a function of item strength.*—In the analyses just presented, the measures employed represented an average of each $S$'s performance on 12 items. It is assumed that such scores indicate what is going on in the learning and recall of individual items. However, few analyses in terms of individual items have been reported (Runquist, 1957). The procedure followed in this experiment, of giving a fixed number of trials on each list, is particularly suited to the analysis of item strength, since an item with a relatively large number of correct anticipations ("reinforcements") is likely to have high strength regardless of whether this strength is the result of its being an easy item or the result of its being learned by a fast learner. This state of affairs should be contrasted with the situation encountered when $S$s learn to criterion. Here, a high number of reinforcements may indicate an easy item or a *slow* learner. Since a slow learner may be assumed to gain less associative strength per reinforcement (Underwood, 1954) there is a confounding of two factors (slow learners and easy items) with opposed effects. Runquist (1957) attempted to reduce this confounding by ranking items within $S$s. Here, no such procedure was felt to be necessary. Not only do the ability of the learner and the ease of the item work in the same way, but an examination of the data revealed that each $S$ covered a considerable range of item strengths, with no overwhelming tendency for any one $S$ to contribute only weak, or only strong, items.

In preparing Fig. 2 and 3, the number of correct anticipations of an item in List 1 learning was taken as a direct measure of that item's List 1 strength, and the successive probability analysis was used to adjust for List 2 strength. All curves were smoothed by the use of three-point moving averages: the 0 and 8 data are based on averages of two points.

Each of the figures presents recall data as a function of the strength of corresponding items on List 1. A corresponding item, for most groups, was the item with the same stimulus. For Cond. $S_N R_N$, the correspondence is formal, and reflects (*a*) minor differences in word frequency, minimized for corresponding words, and (*b*) common relative serial positions in the four presentation orders.
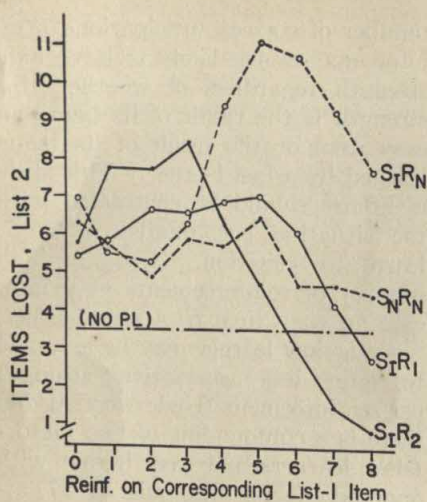
FIG. 2.   Recall losses after 48 hr. as a function of List 1 strength.

Figure 2 shows the losses (obtained recall subtracted from the prediction of the probability analysis) in terms of the *number* of items lost.   Figure 3 shows percentages of recall, obtained by dividing observed by predicted recall and multiplying by 100.   There are no a priori reasons for supposing one way of presenting the data to be better than the other.   However, Underwood[3] has shown that in at least one instance, percentage of recall was monotonically related to strength, as expected, while the relationship between loss and strength was non-monotonic.

Examination of Fig. 2 and 3 reveals that, for some or all degrees of List 1 strength, there is an indication of PI in every group.   In Cond. $S_I R_N$, there appears to be a relative maximum of interference in the vicinity of six reinforcements on List 1.   This is in agreement with Underwood's (1945) finding that maximal interference seemed to result when items were of roughly equal strength on the two lists: assuming a normal distribution of item strengths, it would not be

unreasonable to expect that most item pairs of roughly equal strength would also be of intermediate strength.   The finding is also similar to the familiar fact that, in RI, increased degrees of IL first increase, and then decrease the amount of RI.   Particularly interesting is the high degree of forgetting in the groups with similar responses, when the List 1 response is weak.   This may help to explain why relatively little PI has been obtained with similar responses: it may be necessary to keep List 1 strength low to obtain maximal interference. There is no reason to expect that the finding of maximal interference with approximate equality of the two lists should be generalizable beyond the $S_I R_N$ case.

## DISCUSSION

It will be recalled that Barnes and Underwood (1959) suggested that the unlearning-and-recovery sequence of events applied only to the $S_I R_N$ condition, and that if one assumed no other mechanisms of interference, only $S_I R_N$ would produce PI.   However, it was also proposed that with similar responses, loss of differentiation of list membership might lead to forgetting in a delayed test of recall.   The data on interlist errors in the present experiment, while not provid-
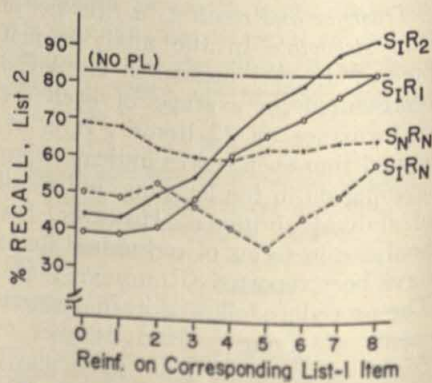


FIG. 3.   Percentage recalled as a function of List 1 strength.

ing a direct test of either mechanism, have an important bearing on this question. All groups with identical stimuli show an increase in interlist errors from 30 sec. to 48 hr. Such errors are rare in Cond. $S_N R_N$, and are not made at all in recall. An increase in the *number* of interlist errors may or may not reflect loss of differentiation, depending on whether one grants primary importance to the error, or to the failure to respond correctly which made an error possible. If one assumes that $S$ first loses the correct response, and then may or may not make an error, then (using the correction for opportunity), the data indicate increased interlist generalization only for $S_I R_N$, the other groups continuing to contribute interlist errors in the same proportion to opportunities at 48 hr. as at 30 sec. If, on the other hand, one assumes that the error displaces a correct response which would otherwise be given, then the appropriate comparison is between the absolute number of errors, which increase in all groups. This would leave unexplained the fact that with similar responses, the increase is proportional to the forgetting obtained, while in $S_I R_N$ it is not. The author's preference is to assume that loss of differentiation is of relatively minor importance, as suggested by the overall equality of total forgetting in all groups except $S_I R_N$, and that the disproportionate increase of errors in the $S_I R_N$ condition is an indication of recovery of List 1 responses. Such a conclusion is in agreement with the implications of Barnes and Underwood's data, and should suffice as a conservative explanation of the present experiment.

Two things remain unexplained: the indication of some PI in the $S_N R_N$ item analysis, and the apparent interaction of PI and List 1 strength in the similar-response conditions. In the case of $S_N R_N$, it is clear that the total stimulus contexts of the two lists are not completely dissimilar—in fact, letters common to the two sets of stimuli may have been sufficient to generate a marginal amount of interlist interference. Other

common stimuli involve the room, $E$, the memory drum, etc. . .

In the similar-response groups, the mediation hypothesis leads one to suspect that the items readily forgotten were those with weak mediators, while the items well recalled were those with strong mediators. Weak mediators might be eliminated by extraexperimental interference, and may provide ideal conditions for the confusion of the two responses. Thus, while loss of differentiation seems not to explain the data as a whole, it may interact with the mediation mechanism, being effective only with low List 1 strength. The explanation of List 2 forgetting as a result of the loss of the List 1 mediator runs into several difficulties. First, the transfer data suggest that when a List 1 item is weak, the dependence of the List 2 response upon it should be minimal—and the loss of an ineffective mediator would be of little consequence. Furthermore, the mediation involved may not be a simple response-chaining process in which the List 1 response serves as a cue for the List 2 response. If a weak List 1 response is strengthened during List 2 learning (as RI experiments seem to suggest), then an interdependence of List 1 and List 2 responses could be built up which would be more complex than the simple response chain assumed to occur when a strong mediator is available. It is obvious that further study of the similar response conditions is necessary before any conclusions can be reached regarding these possibilities.

The data of the present experiment once more indicate the usefulness of considering the unlearning-and-recovery sequence as a basic process in RI and PI, and suggest that while interlist similarity per se is not a highly potent variable, it may yet be found to produce large effects in interaction with degree of learning.

## SUMMARY

Failures to demonstrate an effect of interlist response similarity on PI have previously been reported in experiments in which short retention intervals have been used, with relatively small amounts of PI. In the experiment reported, an attempt was made to

maximize the chances of obtaining PI by using a 48-hr. retention interval.

The design involved 10 groups of *S*s, each of which learned a common List 2 after learning first lists which differed in their similarity to List 2. The first lists had identical stimuli and similar, less similar, or dissimilar responses in three of the basic conditions. Two further conditions involved groups which learned either no List 1, or a List 1 in which both stimuli and responses were dissimilar to those on List 2. In each condition, retention of List 2 was tested after 30 sec. and 48 hr., with independent groups tested at each time interval.

The results showed significant PI only for the condition in which the two lists had identical stimuli and dissimilar responses. However the degree of List 1 strength associated with maximal interference was different for each condition, suggesting that significant PI might be obtained in each condition by appropriate manipulations of List 1 strength.

## REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, 58, 97–105.

BESCH, N. F., & REYNOLDS, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, 55, 554–558.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, 30, 185–200.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, 40, 592–603.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

RUNQUIST, W. N. Retention of verbal associates as a function of strength. *J. exp. Psychol.*, 1957, 54, 369–374.

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, 57, 449–475.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1957.

SPIKER, C. C., & HOLTON, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.*, 1958, 56, 123–132.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teacher's College, Columbia University, 1944.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, 59(3, Whole No. 273).

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, 51, 276–282.

YOUNG, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, 50, 113–119.

# EFFECTS OF NONREINFORCED TRIALS IN TWO-CHOICE LEARNING WITH NONCONTINGENT REINFORCEMENT [1]

JAMES G. GREENO [2]

*University of Minnesota*

The purpose of this paper is to examine the role of nonreinforced trials in a simple prediction situation. Stimulus sampling theories have successfully accounted for many of the results from choice experiments in which some one of the alternative responses is reinforced on each trial (see Estes, 1959). However, these theories have not yet been extended so as to provide an adequate account of results from experiments which include trials on which no reinforcement is presented.

In the present studies, $S$ was instructed to predict which of two lights would flash on each of a series of trials. Following $S$'s choice on each trial, one of three events occurred. On some trials, one of the lights flashed ($E_1$ or $E_2$) constituting a reinforcement for the response of predicting the light that flashed ($A_1$ or $A_2$). On other trials, neither light flashed ($E_0$) constituting a nonreinforced trial. The proportions of trials on which $E_1$, $E_2$, and $E_0$ occurred are denoted $\pi_1$, $\pi_2$, and $\pi_0$, respectively. In this paper, $p_1$ will designate the observed proportion of $A_1$ choices.

Two hypotheses have been offered frequently to account for the effects of $E_0$ trials. One of these is the identity hypothesis, which asserts that an $E_0$ trial leaves choice probabilities unchanged. A second suggestion is the correction hypothesis, which implies that an $E_0$ trial reduces the probability of the response chosen by $S$ on that trial. Estes (1959) summarized the results of several experiments involving $E_0$ trials and suggested that the correction and identity hypotheses describe processes which occur in different degrees in different situations, depending upon such variables as instructions and preliminary training.

In particular, results obtained by Atkinson (1956), Neimark (1953), and Millward (1960) are consistent with identity-hypothesis predictions. However, Anderson and Grant (1957, 1958) and LaBerge, Greeno, and Peterson (1962) have obtained results indicating that with $\pi_1 > \pi_2$, $E_0$ trials reduce $p_1$; and sequential statistics reported in these studies suggest that the obtained changes in $p_1$ may not have been produced by a correction effect.

In most studies of the effects of $E_0$ trials, investigators have tested for quantitative agreement between data and predictions from specific theories. On the other hand, the present study is intended to provide evidence regarding the qualitative properties of the effect of $E_0$ trials. Therefore, $E_0$ trials were presented in situations for which ordinal predictions could be derived so as to differentiate between

hypotheses. In terms of empirical variables, the present experiments were not designed primarily to test the quantitative effects of experimental operations. Rather, these data were obtained in order to determine whether certain variables are relevant in relation to the effects of $E_0$ trials.

## EXPERIMENT I

This study was designed to investigate further the finding that $E_0$ trials reduce $p_1$. Evidence was sought regarding two questions:

1. Is the effect of an $E_0$ trial invariant with respect to the number of trials on which $S$ has received $E_0$? This question arises from the possibility that the effect of $E_0$ trials might decrease over a series of partially reinforced trials. This might occur if $E_0$ events acquired secondary reinforcing properties through association with reinforced trials in the sequence (Bush, 1960), or if the effect of $E_0$ trials depended upon disrupting $S$'s behavior and disruption effects diminished as $E_0$ events continued to occur (Neimark, 1953).

2. Do $E_0$ trials reduce the asymptotic value of $p_1$? LaBerge, Greeno, and Peterson's (1962) results included differences among mean choice frequencies due to $E_0$ trials during acquisition. This second question then, simply asks whether such a difference also occurs during near-asymptotic performance.

## Method

The 72 $S$s were students in introductory psychology classes. Two $S$s at a time were seated in a darkened room and instructed to predict which of two red lights would flash each time they heard a buzzer signal. Each $S$ indicated his prediction on each trial by pressing one of two spring-release levers. The $S$s received eight practice trials: $E_1E_1E_2E_2E_0E_0E_1E_2$. The $E$ emphasized that $S$ "should make a choice on each

trial, no matter what happens," and then asked for questions. If $S$ asked or remarked about the $E_0$ events, $E$ said, "That may happen on some trials. In any case you should make a choice on each trial." No other instructions were given regarding $E_0$ trials. Following the instructions, trials were presented without interruption until all trials had been presented.

Each experimental trial consisted of the following events: buzzer signal, 2 sec.; off, 1 sec.; reinforcement event, 1 sec.; off, 2 sec. This temporal sequence was automatically controlled. On each trial, $S$'s choice and the reinforcement event were automatically recorded.

Reinforcement schedules were constructed by randomly ordering 20-trial blocks of $E_1$ and $E_2$ trials, and then randomly adding the number of $E_0$ trials needed to satisfy the specified value of $\pi_0$. Each block, then, included 20 reinforced trials, and a total of $20/(1 - \pi_0)$ trials.

The three experimental groups received reinforcement sequences as follows:

Group 0—
Blocks 1–10: $\pi_1 : \pi_2 = 90 : 10$; $\pi_0 = 0$.
Group 50—
Blocks 1–10: $\pi_1 : \pi_2 = 90 : 10$; $\pi_0 = .50$.
Group 50$P$—
Blocks $P_1$–$P_3$: $\pi_1 : \pi_2 = 50 : 50$; $\pi_0 = .67$,
Blocks 1–5: $\pi_1 : \pi_2 = 90 : 10$; $\pi_0 = .50$.

A comparison between Groups 50 and 50$P$ is relevant to the question of invariance of the effects of $E_0$ trials. The pretraining schedule received by Group 50$P$ was selected because LaBerge, Greeno, and Peterson (1962) found that this sequence produced no change in $p_1$ for a 90:10 sequence with $\pi_0 = 0$. Thus, if Groups 50 and 50$P$ were to differ over Blocks 1–5 of the present study, then this difference could be attributed to a change in the reinforcing effect of $E_0$ trials due to their presence in the pretraining sequence.

A comparison of Groups 0 and 50 over Blocks 6–10 is relevant to the question regarding near-asymptotic properties of $E_0$ effects.

Six different random schedules were used, with orders of $E_1$ and $E_2$ events matched across groups for comparable blocks. The right- and left-hand response, respectively, was designated $A_1$ for one-half of the $S$s in each condition.

## Results and Discussion

The mean proportions of $A_1$ choices ($p_1$) for each group are graphed by

blocks in Fig. 1. Over the first five blocks of 90:10 reinforcement, $p_1$ was obtained for each $S$. An estimate of $\sigma^2 = .0106$ ($df = 36$) was obtained as the residual mean square of the $3 \times 2 \times 6$ factorial analysis of variance involving experimental groups, right or left sides, and schedules as the factors. Then 90% confidence intervals for orthogonal contrasts between group means were estimated as follows:

$$p_1(0) - \frac{p_1(50) + p_1(50P)}{2} = .091 \pm .052;$$

$$\bar{p}_1(50P) - p_1(50) = -.018 \pm .060.$$

In qualitative terms, $p_1$ was less with $\pi_0 = .50$ than with $\pi_0 = 0$; and $p_1$ with $\pi_0 = .50$ did not differ significantly as a result of the 50:50 pretraining trials. The difference due to $E_0$ trials over these blocks replicates earlier findings and thus provides additional contraindication for the identity hypothesis. Since there was not a significant increase in $p_1$ due to the 50:50 pretraining, there is no evidence that the effect of $E_0$ trials decreased over the partially reinforced sequence of 50:50 trials.

In order to obtain evidence as to whether $E_0$ trials influenced asymptotic values of $p_1$, the value of $p_1$ was obtained for each $S$ in Groups 0 and 50 for each block during Blocks 6–10. These scores were subjected to analysis of trends (Grant, 1956). The $F$ for overall linear trend (1.36; $df = 4/184$; $P > .25$) and the $F$ for difference between group linear trends (1.16; $df = 1/46$; $P > .75$) were not significant. Therefore, there is no statistical evidence that Groups 0 and 50 were not at asymptote during Blocks 6–10. The $F$ for difference between group means was significant (9.28; $df = 1/46$; $P < .0005$), indicating that the asymptotic value of
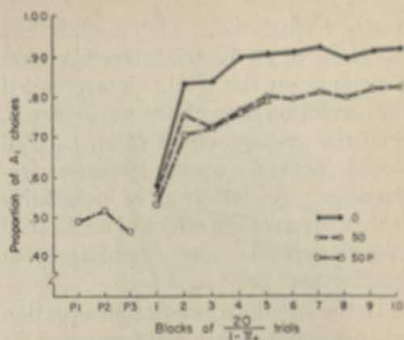


FIG. 1. Mean proportions of $A_1$ choices across blocks of trials in Exp. I. (In Blocks P1–P3, $\pi_1 : \pi_2 = 50:50$. In Blocks 1–10, $\pi_1 : \pi_2 = 90:10$.)

$p_1$ was changed by the $E_0$ trials received by Group 50.

In Table 1 are presented estimated means and variances of $p_1$ scores, and proportions of events ($p_{j,a}$) such that $A_i$ and $E_j$ occurred on Trial $n$, and $A_i$ occurred on Trial $n + 1$. Lines 1 and 2 of Table 1 are relevant to the present discussion. Anderson and Grant (1958) used a statistic based on similar data to estimate changes in $p_1$ across $E_0$ trials. Let $\gamma = p_{0,21}/\pi_0 - p_{0,12}/\pi_0$. For Group 50 of the present study, $\gamma = -.034$, indicating that $p_1$ decreased across $E_0$ trials. This result is consistent with those obtained in Anderson and Grant's analysis, and suggests that the change in $\bar{p}_1$ reported here was due to effects of $E_0$ trials on choice probabilities, rather than to changes in the effects of $E_1$ and $E_2$ events.

Taken together, the obtained difference in $\bar{p}_1$ and the negative value of $\gamma$ provide evidence against the identity hypothesis which is particularly compelling, since it is based on near-asymptotic performance.

## EXPERIMENT II

This experiment was designed to provide evidence relevant to the cor-

rection hypothesis, which implies that the effect of an $E_0$ trial depends upon $S$'s choice on the trial. Therefore, if the correction hypothesis were correct, then the average effect of an $E_0$ trial would depend upon response frequencies. An alternative possibility is that the average effect of an $E_0$ trial might depend upon reinforcement frequencies.

In the standard prediction situation with noncontingent reinforcement, it is impossible to discriminate between the effect of these variables on the basis of ordinal hypotheses, since response and reinforcement frequencies turn out to be equal at asymptote. In order to remove this equality, unequal incentives were introduced for the two response alternatives. Under these conditions, $S$ should choose the response associated with the higher incentive more frequently than that response is reinforced. Thus, it was hoped that the contributions of response frequencies and reinforcement frequencies to the effect of $E_0$ trials could be separated.

Optimal conditions for this purpose would include a condition for which $p_1 = .50$ when $\pi_1 > \pi_2$ and $\pi_0 = 0$, and another condition for which $p_1 < .50$ with $\pi_1 = \pi_2$ and $\pi_0 = 0$. The simplest form of the correction hypothesis would predict that $E_0$ trials should increase $p_1$ in the second condition, and that $E_0$ trials would have no effect in the first condition. On the other hand, if the effect of $E_0$ trials depended upon reinforcement frequencies, there should be no effect due to $E_0$ trials in the second condition, and $p_1$ should be decreased by $E_0$ trials in the first condition.

The incentive operation used was to instruct $S$ that he would receive more points for a correct prediction of one light than the other. Preliminary studies indicated that the optimal conditions described above would be best approximated with a point ratio of 1:6.

## Method

The 250 $Ss$ were students in introductory psychology classes. Procedures and apparatus were the same as those used in Exp. I with the following exceptions: Instead of facing a box with red lights and levers, $S$ faced a slanting panel which held two spring-release buttons with which he indicated his choices. The signal light above the left button was red, and the light above the right button was white. The numerals 1 and 6 were displayed beside the left and right buttons, respectively.

Instructions indicated that the experiment was a test of $S$'s skill at making choices. The task for $S$ was to get as many points as he could by predicting which light would flash on each trial; and $S$ was told that he would receive one point each time he correctly predicted the red left-hand light and six points each time he correctly predicted the white right-hand light. Practice trials and remarks about $E_0$ trials were as in Exp. I.

This experiment consisted of three sections which were run separately. In Section 75, $\pi_1:\pi_2 = 75:25$; in Section 50, $\pi_1:\pi_2 = 50:50$; in Section 25, $\pi_1:\pi_2 = 25:75$. In each section, a group with $\pi_0 = 0$ (Groups 75/0, 50/0, 25/0) was compared with a group with $\pi_0 = .50$ (Groups 75/50, 50/50, 25/50).

In all cases, six blocks of trials were presented. Eight different random schedules were used. Schedules were constructed in the manner described for Exp. I. Schedules were matched within sections with respect to the order of $E_1$ and $E_2$ events. Schedules were matched across sections with respect to the trial numbers on which $E_0$ occurred for the groups with $\pi_0 = .50$. The left-hand response was designated $A_1$ for all $Ss$.

For Sections 75 and 50, a double sampling technique was used to determine the number of $Ss$ to be used (Cox, 1958). In each case, $n$ was set so that the difference between mean values of $p_1$ for $\pi_0$ conditions could be estimated with a 90% confidence interval of length less than .10. The number of $Ss$ in the six experimental groups were as follows: Group 75/0, 64; Group 75/50, 32; Groups 50/0 and 50/50, 48; Groups 25/0 and 25/50, 32.

## Results and Discussion

Mean values of $p_1$ are graphed by blocks in Fig. 2. The value of $p_1$ over

Blocks 3–6 was obtained for each $S$ and these scores were analyzed. Estimates of $\sigma^2$ were obtained as residual mean square terms of factorial analyses of variance (see Table 1). Using these estimates, 90% confidence intervals were estimated for the differences between $\pi_0 = 0$ and $\pi_0 = .50$ conditions as follows:

$$\bar{p}_1(75/0) - \bar{p}_1(75/50) = .079 \pm .046;$$
$$\bar{p}_1(50/0) - \bar{p}_1(50/50) = .053 \pm .048;$$
$$\bar{p}_1(25/0) - \bar{p}_1(25/50) = -.100 \pm .050.$$

Lines 3–6 of Table 1 provide more detailed information regarding performance of these groups. The values of the $\gamma$ statistic described above were estimated as follows: Group 75/50, $-.039$; Group 50/50, $+.006$; Group 25/50, $+.030$. These estimates suggest the following: $p_1$ decreased across $E_0$ trials with $\pi_1:\pi_2 = 75:25$; $p_1$ increased across $E_0$ trials with $\pi_1:\pi_2 = 25:75$; and $p_1$ was virtually unchanged by $E_0$ trials with $\pi_1:\pi_2 = 50:50$.
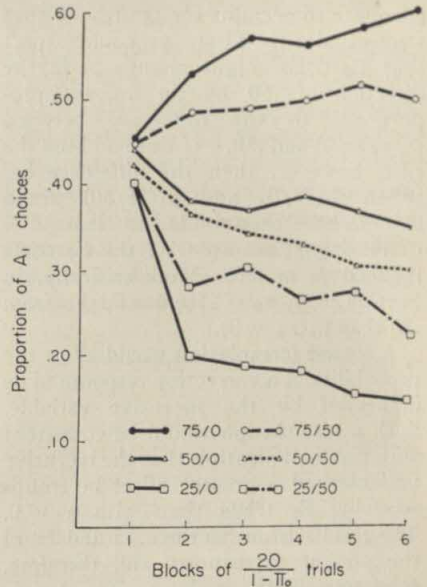


FIG. 2. Proportion of $A_1$ choices across trial blocks in Exp. II.

Two formulations of the correction hypothesis could be applied to these data, although neither of them seems

TABLE 1

ESTIMATES OF GROUP MEANS AND VARIANCES AND PROPORTIONS OF RESPONSE-REINFORCEMENT COMBINATIONS

| Group | $\bar{p}_1$ | $\sigma^2(p_1)$ | $p_{1,11}$ | $p_{1,12}$ | $p_{2,11}$ | $p_{2,12}$ | $p_{0,11}$ | $p_{0,12}$ | $p_{1,21}$ | $p_{1,22}$ | $p_{2,21}$ | $p_{2,22}$ | $p_{0,21}$ | $p_{0,22}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Experiment I, Blocks 6–10 | | | | | | | | |
| 0 | .917 | .0029 | .772 | .052 | .084 | .008 | — | — | .052 | .022 | .008 | .001 | — | — |
| 50 | .814 | | .330 | .038 | .035 | .007 | .334 | .068 | .059 | .023 | .006 | .003 | .051 | .046 |
| | | | | | | Experiment II, Blocks 3–6 | | | | | | | | |
| 75/0 | .584 | .0242 | .310 | .135 | .085 | .056 | — | — | .141 | .165 | .046 | .063 | — | — |
| 75/50 | .505 | .0129 | .138 | .056 | .030 | .030 | .126 | .130 | .079 | .099 | .027 | .037 | .111 | .136 |
| 50/0 | .374 | | .093 | .095 | .082 | .103 | — | — | .112 | .201 | .094 | .221 | — | — |
| | | .0199 | | | | | | | | | | | | |
| 50/50 | .321 | | .047 | .037 | .036 | .047 | .082 | .083 | .043 | .136 | .039 | .128 | .086 | .246 |
| 25/0 | .169 | | .015 | .029 | .024 | .103 | — | — | .045 | .160 | .082 | .543 | — | — |
| | | .0139 | | | | | | | | | | | | |
| 25/50 | .269 | | .013 | .018 | .028 | .076 | .051 | .085 | .026 | .070 | .052 | .220 | .100 | .261 |

Note.—See text for explanations of entries.

adequate to account for the results that were obtained. First, we could expect that $E_0$ trials might change $p_1$ in the direction of .50 by an amount proportional to the difference between $p_1(\pi_0 = 0)$ and .50. If this had been the case, however, then the difference between $\bar{p}_1(75/0)$ and $\bar{p}_1(75/50)$ would seem to have been too large. Instead of decreasing $p_1$ toward .50, the $E_0$ trials reduced $p_1$ to .50. More critically, in Section 50, $\bar{p}_1(\pi_0 = .50)$ was farther from .50 than $\bar{p}_1(\pi_0 = 0)$.

A second formulation would allow the probability of a correction response to be influenced by the incentive variable. Such a formulation would be consistent with results indicating that the incentive variable had a greater effect in groups receiving $E_0$ trials than with $\pi_0 = 0$. The results from Sections 75 and 50 of the present experiment are, therefore, consistent with such a formulation. However, had the incentive operation been more effective in Group 25/50 than Group 25/0, there should not have been a significant difference between these groups in the obtained direction.

In Sections 75 and 50, then, $E_0$ trials reduced the frequency with which $Ss$ chose that response which was reinforced more frequently, although the more frequently reinforced response was associated with different incentives in the two cases. In Section 50, although $\bar{p}_1$ was changed by the presence of $E_0$ trials, it appears that $p_1$ did not change across $E_0$ trials despite the fact that $p_1 < .50$. These results, then, indicate that $S$'s choice frequency is not a relevant variable in relation to the effect of $E_0$ trials; and the present data therefore contraindicate the correction hypothesis. On the other hand, the present findings indicate that the effect of $E_0$ trials is related to the relative frequencies with which the alternative choices are reinforced.

## SUMMARY

Two experiments were conducted in order to obtain evidence regarding the effect of nonreinforced trials in a simple prediction situation. Evidence was obtained regarding two hypotheses: (a) The identity hypothesis, which asserts that an $E_0$ trial does not change choice probabilities; and (b) The correction hypothesis, which implies that an $E_0$ trial reduces the probability of the response chosen by $S$ on that trial.

In Exp. I, two groups receiving sequences of trials including 200 reinforcements with $\pi_1 : \pi_2 = 90:10$ were compared over trial blocks during which $p_1$ was near its asymptote. The group receiving $E_0$ trials showed a lower value of $p_1$ than did the group with $\pi_0 = 0$. This result contraindicates the identity hypothesis. A second comparison from Exp. I examined the effect of a pretraining sequence including $E_0$ trials with $\pi_1 : \pi_2 = 50:50$. The pretraining sequence did not significantly change the effect of $E_0$ trials in the 90:10 sequence which followed. There was, then, no evidence that the effect of $E_0$ trials decreases over sequences of partially reinforced trials.

In Exp. II, unequal incentives were offered $S$ for the two choice responses in an attempt to separate the contributions of choice frequencies and reinforcement frequencies to the effect of $E_0$ trials. It was found that with $\pi_1 : \pi_2 = 75:25$, $E_0$ trials reduced $p_1$; with $\pi_1 : \pi_2 = 25:75$, $E_0$ trials increased $p_1$. With $\pi_1 : \pi_2 = 50:50$, $p_1(\pi_0 = .50)$ was less than $\bar{p}_1(\pi_0 = 0)$, although a sequential statistic indicated that $p_1$ did not change across $E_0$ trials. In each case, $p_1(\pi_0 = 0)$ was less than $\frac{\pi_1}{\pi_1 + \pi_2}$. Therefore, it was concluded that the effect of $E_0$ trials was related to the frequencies with which the choices were reinforced, rather than to the frequencies with which $Ss$ chose the responses. Since the correction hypothesis implies that choice frequencies determine the average effect of $E_0$ trials, the results of Exp. II were interpreted as a disconfirmation of the correction hypothesis.

## REFERENCES

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, 54, 305–317.

ANDERSON, N. H., & GRANT, D. A. Correction and reanalysis. *J. exp. Psychol.*, 1958, 56, 453–454.

ATKINSON, R. D. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, 52, 28–32.

BUSH, R. R. A survey of mathematical learning theory. In R. D. Luce (Ed.),

*Developments in mathematical psychology.* Glencoe, Ill.: Free Press, 1960. Pp. 120–165.

Cox, D. R. *Planning of experiments.* New York: Wiley, 1958.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959. Pp. 380–491.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.,* 1956, **53**, 141–154.

LaBerge, D., Greeno, J. G., & Peterson, O. F. Nonreinforcement and neutraliza-

tion of stimuli. *J. exp. Psychol.,* 1962, 63, 207–213.

Millward, R. B. A comparison of two learning models for two-choice conditioning experiments involving nonreinforced trials. Unpublished doctoral dissertation, Indiana University, 1960.

Neimark, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. Unpublished doctoral dissertation, Indiana University, 1953.

# RETENTION OF FIRST-LIST ASSOCIATIONS AS A FUNCTION OF THE CONDITIONS OF TRANSFER [1]

## LEO POSTMAN

*University of California*

A basic question for theories of transfer and retroactive inhibition (RI) is whether the strength of first-list associations changes systematically during the learning of the second list. A recent study by Barnes and Underwood (1959) has presented clear evidence that the "fate" of first-list associations depends on the conditions of intertask transfer. When the paradigm for negative transfer (identical stimuli and dissimilar responses —A-B, A-C) is used, List 1 associations are unlearned or extinguished during the acquisition of List 2. When the successive tasks conform to the paradigm for positive transfer (identical stimuli and highly similar responses—A-B, A-B'), List 1 associations are maintained at high strength and appear to mediate the reproduction of List 2 responses. The present paper presents additional findings in support of these conclusions.

Experimental evidence consistent with the unlearning hypothesis has been accumulating steadily but has remained short of decisive prior to the study of Barnes and Underwood (Postman, 1961). The hypothesis was first advanced by Melton and Irwin (1940) as part of a two-factor theory of RI. According to this theory, List 1 associations which have been unlearned are not available at the time of recall; others which remain potentially available are displaced by competing associations from List 2. The lack of correlation between total amount of RI and the number of overt interlist intrusions is in accord with this interpretation as is the finding that RI is greater than proactive inhibition (PI) at short retention intervals (Melton & Von Lackum, 1941; Underwood, 1948a). The assumption

that some List 1 associations are not available to S at the time of the retention test received further support from studies using the method of modified free recall or MFR (Briggs, 1954; Briggs, Thompson, & Brogden, 1954; Underwood, 1948b). After learning two successive lists (A-B, A-C), S is presented with the common stimulus term (A) and is required to give either the List 1 response (B) or the List 2 response (C). The relative frequency of List 1 responses declines steadily as a function of the degree of List 2 learning. The fact that the proportion of List 1 responses increases as a function of time supports the interpretation of unlearning as a process akin to extinction followed by spontaneous recovery. The results of MFR tests do not, however, provide crucial evidence for reduced availability of List 1 associations. Since S is instructed to give either B or C, a progressive rise in the proportion of C responses may merely signal increasing dominance of List 2 over List 1 associations and does not compel the conclusion that the latter are not available to S.

A critical test of the unlearning hypothesis requires that the availability of List 1 associations be assessed under conditions in which the effects of response competition and of losses in list differentiation are eliminated. This requirement was met for the first time in the study of Barnes and Underwood (1959). In the experiment using the A-B, A-C paradigm with nonsense syllables as stimuli and adjectives as responses, Ss were presented with the common stimulus terms at the end of List 2 learning and were required to write down both the List 1 and List 2 responses to each of the stimuli. Following a usage adopted by Melton (1961) this modification of MFR will be referred to as MMFR. With List 1 learned to a criterion of one perfect recitation, the MMFR test was administered to different groups after 1, 5, 10, or 20 anticipation trials on List 2. As degree of List 2 learning increased, there was a steady decline in the number of List 1 responses and an equally regular rise in the number of List 2 responses. These trends were obtained regardless of whether credit was given for all responses recalled or only those which were reproduced to the appropriate stimulus and identified

correctly as to list membership. With the exception of the lowest degree of List 2 learning, List 1 responses were reproduced before List 2 responses with only chance or less than chance frequency. These findings give unequivocal support to the unlearning hypothesis and are inconsistent with the "independence hypothesis," according to which the two systems of responses remain independent and intact and RI is due to reproductive inhibition.

The inverse relationship between interlist response similarity and RI suggests that List 1 responses gain in strength during the acquisition of List 2 when intertask transfer is positive. One theoretical account of this relationship assumes that there is generalization of reinforcement between List 1 and List 2 (Osgood, 1946, 1948; Underwood, 1951; Young, 1955). While this hypothesis accounts for the relationship between positive transfer and RI, a second experiment by Barnes and Underwood supports a different conception of the "fate" of List 1 associations under the A-B, A-B' paradigm. In a replication of the procedure described above with lists conforming to the A-B, A-B' paradigm, the MMFR test showed little decline in the recall of List 1 associations as a function of the degree of List 2 learning. Recall of List 2 associations was nearly perfect after only one trial. At the lower degrees of List 2 learning, List 1 responses tended to be recalled before List 2 responses. The total pattern of results, and especially the almost instantaneous acquisition of List 2, points to direct mediation of List 2 responses by List 1 responses (A-B-B') rather than generalization of reinforcement as the mechanism responsible for the positive transfer effects. Since similarity and associative connection between items are correlated, rehearsal of A-B-B' strengthens List 1 responses and at the same time leads to a high level of recall for List 2 responses. Most Ss in the experiment of Barnes and Underwood reported using mediation during List 2 learning.

The study of Barnes and Underwood (1959) represents an important advance in the analysis of the mechanisms of transfer which are fundamental to an interference theory of forgetting. The conclusions concerning the "fate" of List 1 associations will be strengthened if it is possible to show that the differences obtained with the two paradigms are not a function of S's set at the time of recall. Results obtained by the method of anticipation suggest that RI at recall may be enhanced by S's tendency to continue responding from the list which he practiced last. Such

"generalized competition" leads to a large proportion of failures to respond on the test of recall even when the two lists do not share common stimulus terms and List 2 has not been learned to a higher degree than List 1 (Newton & Wickens, 1956; Postman, 1961; Postman & Riley, 1959). It is reasonable to suppose that generalized competition is greater when intertask transfer is negative than when it is positive, especially if there is direct mediation of responses under the latter condition. When recall for homogeneous lists is tested by MMFR, differential effects of set on the reproduction of List 1 responses cannot be ruled out. A mixed-list design was used in the present study in order to assess the differences in the availability of List 1 responses for the A-B, A-C and A-B, A-B' paradigms with the effects of response set equalized. The need for control of response set in MMFR by means of mixed-list designs has been pointed out by Melton (1961).

Performance in MMFR is relatively free from the effects of response competition and list differentiation. A comparison of the amounts of RI as measured by MMFR and the conventional anticipation method will permit an estimate of the extent to which conventional measures of RI reflect reduction in the availability of List 1 responses on the one hand, and response competition and loss of list differentiation on the other. A conventional test of anticipation and MMFR were, therefore, used with different groups to measure RI under the conditions of the present experiment.

## METHOD

*Experimental design.*—Four groups of Ss— two Work groups and two Rest groups— learned a list of eight paired associates (A-B) to a criterion of one perfect recitation. The Work groups were then given 20 trials on a second list of eight paired associates. For half of the pairs in List 2 the relationship to the pairs in List 1 conformed to the A-B, A-C paradigm, and for the other half of the pairs to the A-B, A-B' paradigm. The Rest groups rated a series of pictures on several evaluative dimensions for a period equal to that spent in List 2 learning by the Work groups. At the end of the retention interval one Work group

and one Rest group were given an MMFR test in which the common stimulus terms (A) were presented and Ss were required to give both List 1 and List 2 responses. Following the MMFR test List 1 was relearned for 10 trials. The other Work group and Rest group relearned List 1 for 10 trials without an intervening MMFR test. The four conditions included in the design will be designated as MMFR Work, MMFR Rest, Conventional Work, and Conventional Rest.

*Lists.*—The stimulus terms were eight nonsense syllables of 47–53% association value (Glaze, 1928). The intrastimulus similarity was low. None of the consonants were duplicated, and each of four vowels was repeated once. The pool of response terms consisted of two-syllable adjectives from Haagen's (1949) tables. There were eight sets of three adjectives each. Two of the three adjectives in each set were those used by Barnes and Underwood (1959) as responses in A-B, A-B' pairs and had similarity ratings from .9 to 1.4 on Haagen's scale. The third adjective in each set had no apparent relationship to the other two. Pairs were assigned to lists so that (*a*) each nonsense syllable was the common stimulus for similar responses half the time and for dissimilar responses the other half of the time, and (*b*) similar responses and dissimilar responses from a given set of adjectives were each paired with a common stimulus half the time. There were two combinations of lists; the two lists within each combination were learned first and second equally often. Intralist response similarity was low throughout: there were no duplications of first letters, and no more than one duplication of a terminal suffix.

*Procedure.*—The lists were presented on a Hull-type memory drum at a 2:2 rate, with an 8-sec. intertrial interval. There were four different orders of pairs each of which was used as a starting order equally often. The two lists learned by the Work groups were separated by 2 min. The total retention interval, which was filled by the picture-rating task for the Rest groups, was 15.3 min.

For the MMFR test the drum was operated manually and the exposure of successive stimulus terms was paced by S. The Ss were instructed to call out the two responses in the order in which they occurred to them when the common stimulus term appeared in the window.

*Subjects.*—With two conditions (Work vs. Rest) and two tests of recall (MMFR vs. Conventional), there were four groups of 16 Ss each. The Ss were undergraduate students who were not necessarily naive to rote-

learning experiments but had no previous experience with MMFR tests. For purposes of assignment to conditions, 64 entries were made so that to each S in a Work group there corresponded an S in the appropriate Rest group who learned the same test list with the same starting order. The Ss were run in blocks of 4, with 1 S per block drawn at random from each of the four conditions. The running order within blocks was determined by a table of random numbers. No Ss were lost because of failure to learn.

## RESULTS

*List 1 learning.*—The mean number of trials to criterion on List 1 for the combined groups was 13.26, with an SD of 7.98. The means for individual groups ranged between 12.81 and 14.00 and did not differ significantly ($F < 1$).

The mean number of trials to a criterion of 4/4 for the pairs used in the A-B, A-B' paradigm was 10.39, and 10.31 for the pairs used in the A-B, A-C paradigm. For purposes of this comparison as well as in subsequent analyses of recall and relearning, the classification of the pairs learned by Ss in the Work groups was applied to the protocols of corresponding Ss in the Rest groups.

*List 2 learning.*—During practice on List 2, the A-B' paradigm produced faster learning than the A-C paradigm. The mean number of correct responses on A-B' pairs was 65.06 ($SD = 12.88$), and 59.47 ($SD = 11.15$) on A-C pairs. The results for the Work groups were quite similar—64.94 vs. 59.31, and 65.19 vs. 59.62 for the MMFR and Conventional groups, respectively. The difference between the two kinds of pairs is significant ($t = 2.45$, $df = 31$, $.01 < P < .02$) and is in accord with the differential transfer effects normally predicted for the two paradigms.

The number of interlist intrusions given to the appropriate stimulus terms was greater for the A-B, A-B'

## TABLE 1

NUMBER OF LIST 1 RESPONSES CORRECTLY REPRODUCED IN RECALL AND RELEARNING

| Measure | Cond. | Paradigm | | | |
|---|---|---|---|---|---|
| | | A-B, A-B'[a] | | A-B, A-C[a] | |
| | | Mean | SD | Mean | SD |
| MMFR Groups | | | | | |
| MMFR | Work | 3.06 | 1.20 | 2.12 | .86 |
| | Rest | 3.81 | .39 | 3.75 | .39 |
| Conventional recall | Work | 2.31 | 1.53 | 1.69 | .76 |
| | Rest | 3.43 | .66 | 3.69 | .66 |
| 10 trials of RL | Work | 33.12 | 5.85 | 31.75 | 6.32 |
| | Rest | 37.88 | 2.15 | 37.31 | 2.72 |
| Conventional Groups | | | | | |
| Conventional recall | Work | 2.50 | 1.05 | 1.56 | 1.00 |
| | Rest | 2.94 | .50 | 3.12 | .93 |
| 10 trials of RL | Work | 35.19 | 3.75 | 32.94 | 3.69 |
| | Rest | 36.31 | 2.81 | 36.75 | 4.07 |

[a] Pairs learned by Ss in the Rest groups were classified according to treatment for corresponding Ss in the Work groups.

than for the A-B, A-C paradigm. Twelve Ss contributed 35 intrusions on A-B' pairs whereas 7 Ss made 11 such errors on A-C pairs. This difference agrees with that found in other investigations (Barnes & Underwood, 1959; Underwood, 1951; Young, 1955).

*Transfer from List 1 to List 2.*—To estimate the net amount of transfer from List 1 to List 2, the total numbers of correct responses on the first five trials of acquisition of the two lists were used. Since the fastest learner on List 1 required five trials to reach criterion, all Ss could be included in this analysis. For the A-B, A-B' paradigm the mean number of correct responses rose from 7.50 to 12.22; for A-B, A-C there was only a small increase from 6.75 to 8.12. While the net transfer effect is positive in both cases, the gain is substantially larger for A-B' than for A-C. The

interaction, Pairs × Lists, is significant beyond the .001 level ($F = 14.73$, $df = 1/31$). In the absence of appropriate controls the effects of learning-to-learn and warm up cannot be separated from the specific transfer effects falling on the two classes of pairs. The large difference in the net amount of gain indicates, however, that the specific transfer effects were positive for A-B' and negative for A-C.

*MMFR test.*—The mean numbers of correct responses on the MMFR test are shown in Table 1. The scores are based on responses given to the appropriate stimulus terms. For the Rest group there is only a negligible decline in the recall of List 1. The Work group shows losses for both kinds of List 1 pairs but the amount forgotten is clearly greater after interpolation of A-C than of A-B'. Recall

of List 2 was nearly perfect—3.94 for A-B' and 3.88 for A-C.

In the statistical analysis of the results, the scores of corresponding $S$s in the Work and Rest groups were treated as paired replicates. Since the direction of the differences was predicted, one-tailed tests of significance were used. The overall amount recalled is significantly greater under Rest than Work ($t = 4.67$, $df = 15$, $P < .001$). For purposes of evaluating the differential effects of the conditions of transfer on the amount of RI, the difference between the recall scores for the two sets of pairs was determined for each $S$. The mean difference score is significantly higher for the Work than the Rest group ($t = 2.56$, $.01 < P < .02$). Thus, the reduction in the availability of List 1 responses is reliably greater for the A-B, A-C than the A-B, A-B' paradigm.

There were only a few scattered instances of List 1 responses given to an incorrect stimulus on the MMFR test. Inclusion of these responses raises the total number of responses recalled by 2 and 3 for the A-B, A-B' and A-B, A-C conditions, respectively. For the rest group, there is an increase of 2 on each type of pair. Thus the picture remains virtually unchanged whether or not response to the appropriate stimulus is used as a criterion of correct recall.

*Order of recall in MMFR.*—The numbers of cases in which $S$s recalled both responses to a common stimulus term were 48 for A-B, A-B' and 33 for A-B, A-C. The percentages of cases in which List 1 responses were recalled first were 45.8 and 45.4, respectively, i.e., the frequencies are close to chance in both cases. The corresponding percentages obtained by Barnes and Underwood after 20 trials of interpolated learning were 53 and 43.

Since all percentages are near chance, the differences between the two experiments may be considered minor.

*Conventional recall of List 1.*— Table 1 shows the mean numbers of items recalled by the Conventional groups on the paced test of anticipation. The total recall scores of the Work group are significantly lower than those of the Rest group ($t = 3.51$, $P < .01$). The amount of interference is substantially greater for the A-B, A-C than the A-B, A-B' paradigm, and the difference between the two sets of pairs is again significantly larger for the Work than the Rest group ($t = 2.38$, $.01 < P < .02$). There was only one interlist intrusion to an appropriate stimulus term, which was a substitution of B' for B.

*Comparison of MMFR and conventional recall.*—The overall level of recall is significantly higher in MMFR than in conventional anticipation ($F = 10.36$, $df = 1/15$, $P < .01$, after a Freeman-Tukey square-root transformation). However, the amounts of RI are quite similar under the two conditions of testing. The percentages of RI in MMFR are 19.7 for A-B, A-B' and 43.5 for A-B, A-C. On the conventional test the corresponding percentages are 14.9 and 50.0. The Rest-Work differences interact with the method of testing neither for total recall scores nor for the differences between types of pairs.

*Effect of MMFR on conventional recall.*—Since all groups relearned List 1 by the anticipation method, it is possible to evaluate the effects of an MMFR test on subsequent conventional recall. It must be recognized that the time between the last trial of original learning and the first trial of relearning was slightly longer for the MMFR groups than the Conventional groups because of the interpolated MMFR test. However, this differ-

ence is minor relative to the total length of the retention interval. As Table 1 shows, the performance of the MMFR Rest group exceeds that of the Conventional Rest group on the first trial of relearning whereas the differences between the two Work groups are small and not consistent. Whatever opportunity for rehearsal is provided by the MMFR test appears to have a beneficial effect only for the Rest group. However, the resulting increase in RI is not significant as evaluated by the interaction of the Work-Rest differences with the conditions of testing.

The number of appropriate interlist intrusions also shows some increase after the MMFR test. Whereas there was only one intrusion of a B' item for the Conventional Work group, the MMFR Work group gave four B' responses and two C responses on the first relearning trial. The interpolation of an MMFR test appears to reduce the differentiation between lists.

*Relearning.*—The mean numbers of correct responses in 10 trials of relearning are shown in Table 1. The Conventional groups will be considered first. The difference between the Work group and the Rest group is significant ($t = 2.38, .01 < P < .02$), i.e., there is reliable RI when the total performance in relearning is considered. However, the difference between the amounts of RI under the two conditions of interpolation falls short of significance ($t = 1.64$). As Fig. 1 shows, the level of RI for A-B, A-C declines steeply and converges on that for A-B, A-B'.

The overall difference between the Work group and the Rest group is also significant for relearning after MMFR ($t = 3.30, P < .01$). The amount of RI does not differ significantly for the two paradigms ($t = .54$),
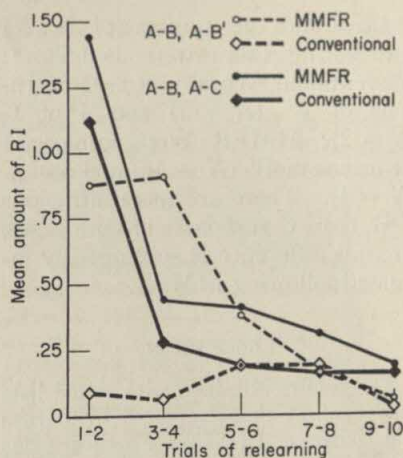


FIG. 1. Amount of RI in relearning as a function of transfer paradigm and condition of testing. (All Ss relearned by the method of anticipation. An MMFR test preceded relearning for the MMFR groups but not for the Conventional groups.)

and the loss scores again converge during relearning (Fig. 1).

Comparison of the temporal trends in RI shows that interpolation of an MMFR test produces a pronounced increase in RI for A-B, A-B' in the early stages of relearning, and a smaller increase for A-B, A-C. As a result, the initial advantage of A-B, A-B' is considerably greater under the conventional treatment than after MMFR: in fact, the relationship between the paradigms is temporarily reversed in the latter case. A trend analysis of the differences between the two sets of pairs yields a significant interaction of Conditions (Work vs. Rest) with the Method of Testing ($F = 4.77, df = 1/15, .02 < P < .05$). Thus, the difference between the slopes of the two RI functions is reduced by the interpolation of an MMFR test.

The effects of MMFR on relearning are reflected in the frequencies of interlist intrusions given to the appropriate stimulus terms. These fre-

quencies and the numbers of $S$s ($N$) contributing them were as follows: Conventional Work group—8 intrusions of B' ($N = 5$) and 4 of C ($N = 2$); MMFR Work group—21 intrusions of B' ($N = 12$) and 5 of C ($N = 4$). There are more intrusions of B' than C responses in both cases, but this difference is substantially increased following MMFR.

## DISCUSSION

The results obtained with mixed lists fully confirm those reported by Barnes and Underwood (1959) for homogeneous lists. The reduction in availability of List 1 associations as measured by MMFR is clearly and significantly greater for the A-B, A-C than the A-B, A-B' paradigm. The relative amounts of interference observed in the two experiments are in substantial agreement. A direct comparison can be made of the percentages that List 1 responses were of all the responses given in MMFR under each condition of interpolation. For A-B, A-B' this percentage was 47.0 in the study of Barnes and Underwood, and 43.8 in the present experiment; for A-B, A-C the corresponding percentages are 35.8 and 35.4.

The consistency of the results obtained with homogeneous and with mixed lists indicates that the differences obtained in MMFR reflect primarily the strength of first-list associations rather than the influence of response sets at the time of recall. This finding parallels that of Twedt and Underwood (1959) that homogeneous lists and mixed lists yield equivalent measures of differential transfer effects.

With response set eliminated as a major determinant of performance in MMFR, the present results give further strong support to the hypothesis that first-list associations are extinguished or unlearned when intertask transfer is negative. The data are also consistent with the hypothesis of response mediation under conditions of positive transfer. When response similarity is high, the

acquisition of List 2 associations is rapid, and List 1 associations are maintained at relatively high strength. However, there is clearly some decline in the strength of A-B after interpolation of A-B'. As Barnes and Underwood have pointed out, the possibility that there is some extinction of A-B during acquisition of A-B' must remain open.

The MMFR procedure is designed to measure associative strength independently of the effects of list differentiation and response competition. Nevertheless the MMFR and Conventional groups yield comparable measures of RI and are equally sensitive to the differences between the paradigms of transfer. It appears that in conventional recall the amounts of RI observed immediately after the end of interpolated learning are largely a function of associative strength rather than degree of list differentiation and response competition. The relationship between the measures of RI obtained in MMFR and conventional recall is likely to change, however, with the progressive decline of list differentiation as a function of time.

The rise in intrusions following MMFR indicates a systematic decrease in list differentiation after a test in which $S$s may give the two responses to a common stimulus term in either order. The fact that the increase in intrusions is greater for A-B, A-B' than for A-B, A-C lends some indirect support to the view that mediation occurs in the former case. The associative connection between highly similar responses may be assumed to be bidirectional, i.e., the pre-experimental probability of B'-B should be equal to that of B-B'. Both sequences occur with approximately equal frequency in MMFR. Reversal of the order of responses in MMFR appears to be conducive to competition between the two alternative sequences and results in persistent interlist intrusions and RI in relearning. There are presumably no such competing sequences in the case of the A-B, A-C paradigm. These effects of MMFR on relearning add to the evidence for the lack of independence between successive tests of retention.

## SUMMARY

This study is concerned with the changes in the strength of first-list associations during the acquisition of a second list. A mixed-list design was used so that for half the syllable-adjective pairs in the two lists the paradigm of transfer was A-B, A-B', and for the other half, A-B, A-C. List 1 was learned to one perfect recitation and List 2 was practiced for 20 trials. To determine the availability of List 1 and List 2 associations at the end of interpolated learning, a procedure devised by Barnes and Underwood (1959) was followed and Ss were required to give both responses to each stimulus in an unpaced test of recall (MMFR). A control group learned and recalled a single list. In a parallel experiment retention for List 1 was tested by the conventional method of anticipation.

In agreement with the results obtained by Barnes and Underwood for homogeneous lists, the MMFR test shows a significantly greater reduction in the availability of List 1 responses for the A-B, A-C than the A-B, A-B' paradigm. Recall of List 2 was nearly perfect. These findings support the hypothesis of unlearning and make it unlikely that the differences in the availability of List 1 associations are a function of response sets characteristic of the two paradigms. The conventional anticipation method yields measures of RI which closely parallel those obtained in MMFR. It is concluded that under both conditions of measurement the amount of RI immediately after interpolated learning is determined largely if not entirely by the availability of List 1 associations.

## REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, 58, 97–105.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, 47, 285–293.

BRIGGS, G. E., THOMPSON, R. F., & BROGDEN, W. J. Retention functions in reproductive inhibition. *J. exp. Psychol.*, 1954, 48, 419–428.

GLAZE, J. A. The association value of nonsense syllables. *J. genet Psychol.*, 1928, 35, 255–269.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association-value ratings for 400 pairs of common adjectives. *J. Psychol.*, 1949, 30, 185–200.

MELTON, A. W. Comments on Professor Postman's paper. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

MELTON, A. W., & IRWIN, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, 53, 175–203.

MELTON, A. W., & VON LACKUM, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition *Amer. J. Psychol.*, 1941, 54, 157–173.

NEWTON, J. M., & WICKENS, D. D. Retroactive inhibition as a function of the temporal position of interpolated learning. *J. exp. Psychol.*, 1956, 51, 149–154.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, 36, 277–301.

OSGOOD, C. E. An investigation into the causes of retroactive inhibition. *J. exp. Psychol.*, 1948, 38, 132–154.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

POSTMAN, L., & RILEY, D. A. Degree of learning and interserial interference in retention. *U. Calif. Publ. Psychol.*, 1959, 8, 271–396.

TWEDT. H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, 58, 111–116.

UNDERWOOD, B. J. Proactive and retroactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, 38, 29–38. (a)

UNDERWOOD, B. J. "Spontaneous" recovery of verbal associations. *J. exp. Psychol.*, 1948, 38, 429–439. (b)

UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first-list learning. *J. exp. Psychol.*, 1951, 42, 44–53.

YOUNG, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, 50, 113–119.

# EFFECTS OF NONREINFORCEMENT ON SUBSEQUENT REINFORCED RUNNING BEHAVIOR [1]

SHELBY J. HARRIS,[2] M. GLENN SMITH,[3] AND SOLOMON WEINSTOCK [4]

*Lehigh University*

Bush and Mosteller (1955, p. 89) present a descriptive model for response acquisition under partial reinforcement which is formally identical with models arising from statistical learning theory. They assume two linear operators on probability which are applied independently, one on reinforced and the other on nonreinforced trials. The model yields the result that asymptotic probability of responding is directly related to the proportion of reinforced trials. This result may also be shown to hold for the asymptotic speed of running in an alleyway. The latter was directly contradicted by the finding by Notterman (1951), Weinstock (1958), and others, that continuously reinforced *S*s reached lower terminal speeds of running than partially reinforced *S*s.

Weinstock (1953, 1958) hypothesized that the effect of nonreinforced trials was to provide an opportunity for nonfunctional response components in the response chain to be eliminated. Elimination of nonfunctional movements should result in an increase in the value of the limit-point of the reinforced trial operator. Weinstock (1953) proposed a modification of the original model in which the limit-point of the reinforced trial operator is assumed to be an increasing function of the number of nonreinforced trials.

In the present experiment two groups of rats were trained to run in an alleyway under 50% and 100% reinforcement, respectively. Half of each group then received extinction. Finally, all *S*s were given 100% reinforcement. In the last phase of the experiment only the reinforced trial operator is applicable and the changes induced in the limit-point should be clearly exhibited.

It was expected that both the partial reinforcement and extinction variables would lead to faster running and that an interaction between the two variables would be found. It was also expected that the 50% group which did not get extinction would run faster when switched to 100% reinforcement since the nonreinforced trial operator would no longer be applied. Finally, it was hoped that the differences induced in speed of running would correspond to differences in within-trial variability. Specifically, it was thought that faster running was due to conditioning of a response chain that contained fewer nonfunctional movements. As *S* traversed the alleyway, the successive placements of its left rear paw were recorded. More general information about the effect of reinforcement schedules on the variability of running behavior was also sought.

## METHOD

*Subjects.*—The *S*s were 40 experimentally naive female albino rats, 105 to 135 days old at the beginning of the experiment.

*Apparatus.*—The apparatus was a straight alley, 10 in. wide and 36 in. long, with a 10 × 18 in. goal box and a start box 5 in. wide

[2] Now at Wells College.
[3] Now at Tufts University.
[4] Now at Brooklyn College.

and 10 in. long. The height of all sections was 8 in. An entranceway, 5 in. wide and 4 in. long, separated the alley from the goal box. Guillotine doors were used in the start box and in the goal box. A 5 × 7 in. vestibule in the far right corner of the goal box contained a drinking tube. On nonreinforced trials a sliding door, which was not visible before S interrupted the last photocell beam, prevented access to the drinking tube. Latency and total running time were measured using Standard Electric .01-sec. timers operated by photocells located 2 in. from the start box and 9 in. inside the goal box, respectively.

Records of the path taken by S were obtained by means of a 20-channel "contact" recorder in which S's left rear paw served to key a thyratron circuit when the paw made contact with the floor. This was accomplished by constructing the floor of tinned copper bus bars embedded in insulated wood at $\frac{1}{4}$-in. intervals. When the paw, which was coated with silver paint, made contact with adjacent bars a thyratron circuit was completed. Resistance in the other paws was too high to complete the circuit.

The runway floor was divided into a checkerboard of $\frac{3}{4}$ × 2 in. sections. Contact within one of these sections activated two electrodes of a 20-electrode paper-burn recorder. One of the electrodes recorded the position of the paw across the runway and the other the position down the length of the runway. Eleven $\frac{3}{4}$-in. and one $1\frac{1}{2}$-in. segments, adjacent to the right hand wall, made up the width of the alleyway. Six 2-in. sections made up the first 12 in. of the length of the runway. Two repetitions of this matrix were used for the remaining portion of the runway proper and a third repetition permitted recording in the goal box. Two photocells, one located 18 in. from the start box and the other 9 in. inside the goal box, operated two of the remaining electrodes of the recorder to aid in determining when S moved from one section to the next.

*Procedure.*—Prior to the experiment, Ss were handled and habituated to 22 hr. water deprivation for 10 days. Throughout the experiment two trials per day with an inter-trial interval of about 45 min. were run under 22 hr. water deprivation. Reinforcement consisted of 10 sec. of drinking. On non-reinforced trials S was confined to the goal box for 30 sec.

Before each trial S's left rear paw was coated with silver paint. After a 30-sec. interval to allow the paint to dry, S was placed in the start box facing the door. Closing the cover depressed a microswitch which activated a 3-sec. thermo-delay relay. When the relay closed a solenoid was energized which served to release a lever which was held by stretched rubber bands and both timers were started. Movement of the lever raised the start-box door.

Two Es (SJH and SW) ran half of the Ss each. The third E coated S's paw with paint and recorded the data.

During the first four trials of the experiment, S was removed from the apparatus if it failed to leave the start box within 20 min. or spent more than 20 min. in the runway or in the goal box without being reinforced. After the fourth trial the removal criteria were reduced to 5 min. Failure to run on three successive trials was used as a criterion for eliminating S from the experiment.

All Ss were reinforced on each of the first 12 trials. On the basis of running speed data from the early trials Ss were assigned to four groups in such a way that the group means were equal.[5] Group CE remained on 100% reinforcement during the acquisition period and was then given 24 extinction trials. Group CC remained on 100% reinforcement, both during acquisition and during the subsequent 24 trials. Group PE received 50% reinforcement during acquisition and then received 24 extinction trials, while Group PP received 50% reinforcement, both during acquisition and the following 24 trials. For Ss receiving 50% reinforcement 6 trials on which reinforcement was to be given were selected randomly from each block of 12 trials.

Mean reciprocal running times were computed for each block of six trials during the course of acquisition. The decision to terminate acquisition was based on inspection of these data. When it appeared Ss' running times were stable, an additional 12 trials were run and the acquisition period terminated. As a result of this procedure a total of 96 acquisition trials were run. Prior to the extinction period all Ss received 2 reinforced trials. A reacquisition period of 30 trials followed the 24 extinction trials. All Ss received 100% reinforcement during reacquisition.

## RESULTS

Latency and running time were converted to reciprocals throughout.

[5] One S was discarded after Trial 9 in accordance with the criteria previously described. Hence, for Group PE $N = 9$, while for the other groups $N = 10$.
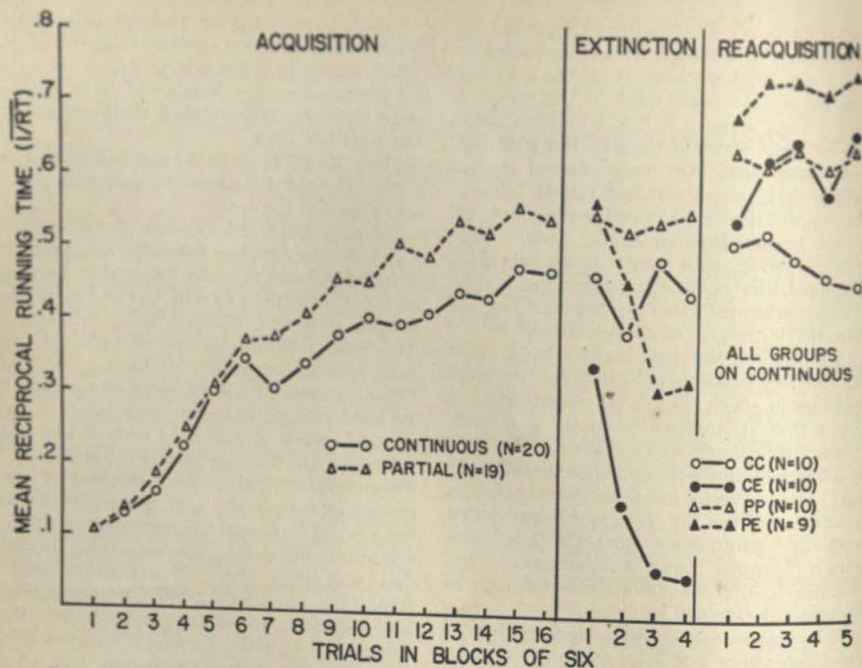
FIG. 1. Mean reciprocal running time plotted as a function of trials in blocks of six.

A record of successive placements of S's left rear paw was obtained on each trial. The paw placements were plotted on a scale drawing of the runway and connected by straight lines yielding an approximate path for each trial. The points of intersection of the path with lines perpendicular to the side walls at distances of 7, 14, 21, and 28 in., respectively, were recorded. Thus, the lateral position of S (i.e., across the width of the runway) at each of the four points down the runway could be estimated. From these data a number of measures involving spatial location could be computed for each S for any block of trials: (a) the variability of lateral position across a block of trials at each of the four points, (b) the mean lateral position across all four points and across trials, (c) the variability of this mean position across trials, (d) the mean across trials of the variance

within a trial computed about the mean position, and (e) the sum within each trial of the absolute deviations from a straight line connecting the start box and goal box and the average of these sums over trials.

In Fig. 1 the reciprocal running time measure is plotted against blocks of six trials for all three phases of the experiment.[6] Reacquisition occurred much more rapidly than original acquisition. Group CE went from a level of responding below that of the first block of acquisition trials to a level about that of Group CC within the first block of reacquisition trials. The major portion of the change in responding for Group CE took place in the first three reacquisition trials.

An analysis of variance for the last 18 trials of acquisition yielded a non-

[6] The two trials at the end of acquisition on which all Ss received reinforcement are omitted from Fig. 1.

significant $F$ ($F = 2.60$, $df = 1/35$, $P > .05$) for the difference between the continuous and partial reinforcement groups. The difference between Groups CC and PP continued to be nonsignificant for the 24 trials of the extinction phase, while the difference between Groups CE and PE yielded an $F$ of 20.32 with $1/15$ $df$, which is significant at the 5% level.

The results of an analysis of variance based on the last 12 trials of the reacquisition phase are summarized in Table 1. Since 7 $F$s were computed, Hartley's $F$ maximum test was used to adjust the experiment-wise error rate to 5%. Pearson and Hartley (1954, Table 19) give the 5% value as slightly less than 8.21. Thus, with the experiment-wise error rate controlled at 5% the effects of the reinforcement and extinction variables would be declared significant while the interaction between $E$s and extinction would be declared nonsignificant. It should be noted that the size of the interaction was due to a larger difference between extinction and no-extinction groups for one $E$ (SJH) than for the other (SW). The direction of the difference was the same for both $E$s.

Group PP showed an increase in speed of running after being shifted to continuous reinforcement.

The group's mean level during the extinction and the reacquisition phases was 0.54 and 0.63, respectively. A $t$ for paired measures, based on the last 18 trials of each of the two phases, yielded a value of 3.04 with 9 $df$, which is significant at the 5% level.

Sixteen analyses employing the measures based on $S$'s path were performed, 8 at the end of acquisition and 8 at the end of reacquisition. With one exception they proved to be nonsignificant. At the end of acquisition the partial reinforcement group showed "significantly" more variability in lateral position at a distance of 14 in. from the start box. While no method of adjusting these results to a 5% experiment-wise error rate exists, common sense suggests that the result not be declared significant.

Mean reciprocal running times for the last 12 trials of initial acquisition were correlated with four measures of $S$s' paths on the same trials. The correlations with running speed were: (a) mean position of $S$ in the lateral dimension (i.e. width) of the runway, $-.09$; (b) average within-trial variance of lateral position computed about the mean position, $-.24$; (c) variance of the mean lateral position across trials, $.03$; and (d) average sum of absolute deviations from a straight line connecting the start and goal boxes, $-.20$.

## TABLE 1
### ANALYSIS OF VARIANCE FOR THE REACQUISITION PHASE

| Source | $df$ | $MS$ | $F$ |
|---|---|---|---|
| $E$s | 1 | .0077 | — |
| Continuous vs. Partial | 1 | .1811 | 12.32* |
| Extinction vs. No-Ext. | 1 | .1819 | 12.37* |
| $E$s × C-P | 1 | .0154 | 1.05 |
| $E$s × E-NE | 1 | .0729 | 4.96* |
| E-NE × C-P | 1 | .0097 | — |
| E-NE × C-P × $E$s | 1 | .0222 | 1.51 |
| Within (error) | 31 | .0147 | |

* $P < .05$.

## DISCUSSION

The major findings involving the time data of this experiment were that non-reinforced trials, administered either in an extinction or partial reinforcement procedure, led to a change in the terminal level of responding under subsequent continuous reinforcement. There was no sign of a decrease in the higher level of responding induced by nonreinforcement, which suggests that a nonreversible change in the reinforced trial operator has occurred. Although appropriate

controls were lacking, the results also suggest that the larger the number of nonreinforced trials, the greater the effect induced on the reinforced trial operator. To this extent the qualitative predictions stemming from a modification of the Bush and Mosteller model received support. It is difficult to assess the failure to find the predicted interaction between the extinction and proportion of reinforcement variables. Ideally a power analysis should be performed. However, this is impossible without a good estimate of the expected magnitude of the interaction, which depends critically on the unknown parameters of the various learning operators.

The problem of providing a theoretical underpining for the induced changes in the reinforced trial operator remains. One attempt in this direction (Weinstock, 1953, 1958) leads to the prediction that the changes in speed of running should be reflected in a reduction of within-trial variability. This prediction received no support from the present results.

In general, the spatial location measures were not related either to the independent variables investigated or to speed of running. The question of the reliability of the measures arises. To obtain a rough answer, values for each of the four major measures for Trials 86 to 91 were correlated with those for Trials 92 to 97. Values of .77, .66, .86, and −.19 were obtained for the absolute deviations from the midline, the variance within a trial, the mean position, and the variance of the mean position, respectively. For purposes of comparison the correlation coefficient was also computed for the running time measure for the same blocks of trials and a value of .83 was obtained. With the exception of the variance of the mean position, it seems that the measures taken were sufficiently reliable.

## SUMMARY

Forty albino rats were given two trials per day, separated by 45 min., in a straight alley under 22 hr. of water deprivation. Half of the Ss received 100% and half 50% reinforcement for 96 trials of acquisition. The two groups were further halved with half of each receiving 24 extinction trials and the other half continuing on their reinforcement schedules. All Ss then received 30 trials under 100% reinforcement. In addition to latency and running time measures, detailed recording was made, by use of a special "contact" recorder, of the path taken by S.

The following results were obtained. At the end of acquisition 50% Ss were running more rapidly than 100% Ss but the difference was not statistically significant. During the last 30 trials of the experiment Ss who had had 50% reinforcement were running more rapidly than Ss who had had 100% reinforcement, and Ss who had had extinction were running more rapidly than those who had not. The 50% reinforcement Ss ran more rapidly after being shifted to 100%. These differences were statistically significant. The interaction between the extinction and proportion of reinforcement variables was not significant.

In general the various measures of spatial location and variability which were taken showed no relation to the independent variables and quite low correlations with speed of running.

## REFERENCES

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* N. Y.: Wiley, 1955.

NOTTERMAN, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *J. exp. Psychol.*, 1951, 42, 273–291.

PEARSON, E. S., & HARTLEY, H. O. *Biometrika tables for statisticians.* Cambridge: Cambridge Univer. Press, 1954.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. Unpublished doctoral dissertation, Indiana University, 1953.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, 56, 151–158.

# EFFECTS OF ORIENTING TASK, PRACTICE, AND INCENTIVE ON SIMULTANEOUS INCIDENTAL AND INTENTIONAL LEARNING [1]

## ARNOLD MECHANIC [2]

### *Northwestern University*

In the Type I incidental learning situation, S is given no instructions to learn but is later tested for the materials to which he was exposed. In the Type II situation, S is exposed to two sets of materials, instructed to learn only one of the sets, and is later tested for the materials which he was not instructed to learn. The finding that there is less incidental learning in the Type II situation than in Type I has been attributed to competition from the intentional task for the limited amount of exposure time which is available to S (Mechanic, 1962). It could therefore be expected that certain intentional learning variables would adversely affect Type II incidental learning. Any experimental variation which focuses S on the intentional items should serve to reduce the proportion of the total exposure time available for responding to the incidental items. Examples of such variables would be strength of the incentive motivating S to learn, and difficulty of the intentional items.

Bahrick (1954) and Bahrick, Fitts, and Rankin (1952), using nonverbal materials, found that the amount of learning irrelevant to a set is inversely related to the strength of the incentive determining that set. Mechanic

(1962), however, found that incidental learning of verbal items did not vary as a function of the difficulty of the concurrently learned intentional items. The divergence between these findings, with regard to the task-competition hypothesis, becomes explicable when we consider the different orienting tasks which Ss were required to perform. Mechanic's Ss were required to pronounce the incidental items in order to rate phonetic similarity. Thus, competition from the intentional items (for the available exposure time) could not prevent Ss from responding to the incidental items. In Bahrick's study, on the other hand, there was no specific orienting task which required S to respond to the incidental stimuli. The S could, for example, learn geometric forms intentionally without necessarily responding to their incidental colors, and large amounts of competition from the intentional task could result in a failure to respond to the incidental stimuli.

If the above explanation is appropriate, it would be expected that the incentive for intentional learning would have little or no effect on concurrent incidental learning when the orienting task required that S respond to the incidental stimuli. This experiment is an attempt to clarify the joint effects of orienting task and intentional learning variables on Type II incidental learning. By varying the degree of responding to the incidental items required by the orienting task,

it should be possible to get a clearer picture of how intentional learning variables influence Type II incidental learning. Because the intentional learning variable that has been chosen is incentive, it will, at the same time, be possible to assess the effects of incentive upon the intentional rote learning of verbal materials.

## METHOD

*General design.*—Three orienting tasks were selected so as to differ in the degree of responding that they required to the incidental stimuli. Two levels of incentive were used. In addition, two degrees of practice (number of presentations) were used in order to obtain greater generality in the evaluation of the incentive and orienting task variables, and in order to study possible interactions between these variables and practice. Twelve groups of 20 Ss each were used in a standard $3 \times 2 \times 2$ factorial design.

All Ss were given a list made up of 12 pairs of trigrams. They were instructed to learn one member from each pair and were discouraged from attempting to learn the other member of the pair. Different groups of Ss were exposed to the items by means of the three orienting tasks described below. The Ss were told that the two members of each pair were syllables from two different primitive languages. The first members of each pair were said to come from one language while the second members of the pairs came from the other language. In addition to the orienting task instructions, Ss were told that we also wanted to study the ease of learning of one primitive language under conditions of distraction from a second language. In this way, instructions to learn were given for only one member of each pair (intentional learning of one member and incidental learning of the other). After the list presentations, a test of free recall was given for all the items. Each S received two learning scores, one being the number of correctly reproduced intentional items and the other being the number of correctly reproduced incidental items.

*Orienting tasks.*—Three orienting tasks were devised so as to produce differing numbers of "pronouncing" responses to the incidental items. By a pronouncing response, reference is made to a hypothetical response by S so that he reacts to the trigram as a single pronunciable unit. The operations involved in the three orienting tasks should

serve to further clarify this definition. The choice of the pronouncing response as an important response for rote learning is consistent with the recent conclusion of Underwood and Schulz (1960) that, "*emitted frequency is of fundamental importance in the integration of response elements* [italics mine]" (p. 291). The orienting task devised to produce maximal responding to the incidental items will be called Phonetic Similarity ratings (PS). To PS Ss, the experiment was introduced as a test of the notion that words meaning the same thing in different languages tend to sound more alike than words with different meanings. The Ss were asked to judge the phonetic similarity between the members of each pair of trigrams. The Ss were required to pronounce the members of each pair to themselves and to rate their phonetic similarity on a five-point scale. Following these instructions, learning instructions for only one set of items were given as described above.

It may be noted that this orienting task required that S respond to the incidental items as single pronunciable units. Another orienting task was designed in order to minimize the probability of S responding to the incidental items in this manner. This minimal responding situation will be referred to as Letter Cancellation (LC). The LC Ss were told that we were interested in the effects of quick changes of set upon letter cancellation performance. They were required to cancel out the vowels on the top language (pair member) and cancel out the consonants on the bottom language. All letters other than A, E, I, O, and U were considered consonants. The Ss were asked to work carefully and were not permitted to erase after having marked a letter. It is clear that this orienting task exposes S to the incidental stimuli without necessarily requiring that he respond to the items as pronunciable units.

Having defined extreme points on a hypothetical dimension, an intermediate orienting task was selected, which will be called Extrasensory Perception (ESP). The ESP Ss were told that of the 144 possible pairings of the items from the two languages, 50 had been selected randomly and printed on cards. These 50 cards were said to be in an envelope that was left on the desk in front of the room. The Ss were told that we wanted to check the contention of previous investigators that people could guess better than would be expected by chance in such a situation. They were required to look at each pair of syllables and then rate the likelihood of its being represented in the envelope. As with the

phonetic similarity judgments, the ratings of likelihood were on a five-point scale. This kind of orienting task was chosen because it seemed to offer some probability of $S$ responding to the incidental items as pronunciable units. At the same time, such responses were not required as they were in the PS situation, and should be of lower frequency for the ESP group. However, the ESP group could clearly be expected to make more pronouncing responses than the LC group.

It should be evident that the amount of time required to perform each of the three orienting tasks will vary, and that effective exposure times for learning will vary concomitantly. Although there are no quantitative data on this issue, it is apparent from observations made during the experiment that PS ratings take the most time while LC takes the least time. Because both intentional and incidental scores were obtained from each $S$, it will be possible to gauge the incidental scores for each orienting task by the corresponding intentional scores for that condition. The intentional scores may be thought of as an index of the time available for learning to take place under each respective orienting task.

*Incentive conditions.*—The $Ss$ were introductory psychology students at Northwestern University who met a course requirement by serving as $Ss$. In addition, "incentive" $Ss$ were offered the possibility of one "bonus hour" toward the required total of 10 hr. They were told that an extra hour of credit would be given to all $Ss$ who learned more of the intentional syllables than the average number learned by the group. It was added that they would be eligible for the bonus only if their performance on the other task (orienting) met certain standards of acceptability. The LC and PS $Ss$ were told that they would be eligible for the bonus only if their performances met minimal standards of accuracy. The ESP $Ss$ were told that they would not be eligible if their guesses were in the bottom 10% of the group. It was made clear, however, that half of the $Ss$ in the group would win bonus hours and would be informed of these awards promptly.

The "nonincentive" $Ss$ were not offered bonus hours, but were otherwise read instructions identical to those of their incentive counterparts. Incentive $Ss$ were questioned about the effects of the incentive during a postexperimental inquiry. Those who stated that they did not need the bonus hour, and that they did not try any harder as a result of the incentive, were discarded and replaced. The positive statements of many $Ss$ would seem to indicate that a bonus hour has incentive value in the same sense as do the small monetary bonuses paid by Bahrick (1954).

*Practice conditions.*—The list of 12 pairs was exposed for either two or five presentations. Successive presentations of the list involved different random orders for the set of 12 intentional items and the set of 12 incidental items. As a result, the pairings of individual items from the two sets were randomly varied from presentation to presentation. This was done to prevent the learning of incidental items through the mediation of the intentional members of the pairs.

*Design details.*—The 12-pair lists were constructed from two separate 12-unit trigram lists. The two 12-unit lists were equated for meaningfulness and were made up of pronunciable high-frequency items. The two lists and their method of construction have been presented elsewhere (Mechanic, 1962). For clarity of exposition, henceforth these 12-unit lists will be referred to as *sets* and the term *list* will refer to the 12-pair lists actually presented to the $Ss$. Each of the 12 pairs in a list was arranged vertically with one member in the top position and the other member directly below it. This arrangement was used to prevent the left to right chaining which might result from the reading habits of $Ss$. The top members of the pairs came from one of the stimulus sets while the bottom members came from the other set. Through presentation of the list during training, both sets of items were presented concurrently to $S$. The $Ss$ were, of course, instructed to learn only one of the two sets.

Five random orders were prepared for each of the sets. This was done to obtain different orders of the items with successive presentations of the list. Additionally this insured that the pairings of individual items from the two sets were randomly varied from trial to trial. For each of the five orders of presentation, two variations were prepared. These reversed the top and bottom member positions but were otherwise identical.

The $Ss$ were given a number of sheets of paper, each covered with a strip of cardboard in which a cut-out window allowed exposure of one pair of syllables at a time in accordance with $E$'s instructions. The pairs of syllables on each sheet were numbered 1 through 12. When $E$ called out the number 1, $S$ moved the window to expose the first pair of syllables and responded to the items in accordance with the requirements of the orienting task. The procedure was repeated for each successive

## TABLE 1
### Average Numbers of Intentional and Incidental Items Recalled by the Various Groups

| Type of Item and Incentive | Two Presentations | | | | | | Five Presentations | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Task LC | | Task ESP | | Task PS | | Task LC | | Task ESP | | Task PS | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Incidental Items | | | | | | | | | | | | |
| Normal | .70 | .56 | 1.90 | 1.38 | 2.75 | 1.48 | 1.85 | 1.77 | 2.65 | 1.93 | 4.50 | 2.17 |
| Augmented | .65 | 1.11 | 1.30 | 1.05 | 2.75 | 2.16 | 1.75 | 1.58 | 2.35 | 2.24 | 4.45 | 1.66 |
| Total | .68 | | 1.60 | | 2.75 | | 1.80 | | 2.50 | | 4.48 | |
| Intentional Items | | | | | | | | | | | | |
| Normal | 7.55 | 2.16 | 6.75 | 2.62 | 5.65 | 1.53 | 10.25 | 1.97 | 8.35 | 1.96 | 8.15 | 2.24 |
| Augmented | 7.40 | 1.50 | 7.05 | 2.16 | 5.80 | 2.18 | 10.30 | 1.95 | 8.90 | 2.30 | 9.20 | 1.91 |
| Total | 7.48 | | 6.90 | | 5.72 | | 10.28 | | 8.62 | | 8.68 | |

pair until all 12 pairs on the page had been exposed. The pairs were presented at a 10-sec. rate, with 20 sec. between presentations of separate sheets. A 5-min. test of free recall was given 1 min. after the end of practice. The Ss were requested to write down as many of the syllables as they could remember regardless of whether they appeared in the top or bottom positions of the pairs. They were told to write the syllables down in the order they were remembered, and that their scores would depend only upon the number of correct items that they wrote down. When the recall test was completed, a postexperimental inquiry was made to ascertain whether S understood the instructions, whether the bonus hour had value as an incentive, and whether an attempt was made to learn the incidental items.

For each of the 12 groups, half the Ss received materials with the items from one set in the top positions of the pairs. For the other half of the group, the otherwise identical pairs were inverted. In addition, half the Ss given each variation of materials were asked to learn the top items and the other half were asked to learn the bottom items.

The Ss were tested in a large room in groups averaging 14 in number. Conditions were assigned to the groups in a random order for each replication of the basic experimental design. Instructions to learn top items were alternated with instructions to learn bottom items for the different groups serving in the same experimental conditions. The Ss within each group were assigned the two different sets of stimulus materials (top-bottom variations) by rotation. The only restrictions on these procedures were to correct for equal Ns.

Because Ss were tested in groups, extra Ss were obtained in most of the experimental variations. Those used in the analyses were selected at random from the total number of Ss in each variation. These selections were made after discarding Ss whose postexperimental reports indicated that they did not understand the instructions, that they tried to learn the incidental items, or that the bonus hour had no value as an incentive. Less than 10% of the Ss appearing for the experiment were discarded for one of these reasons.

## Results and Discussion

Table 1 contains means and SDs for both the incidental and intentional learning scores for the 12 basic conditions. The incidental and intentional scores for any combination of experimental variables were obtained concurrently from the same Ss. Analyses of variance of the incidental and intentional scores were done separately.

*Incentive.*—If incentive is to be used to study the effects of an intentional learning variable on concurrent incidental learning, it should be possible to demonstrate that incentive is an

effective variable with regard to the intentional items. The data in Table 1 indicate that incentive has no significant effect upon intentional rote verbal learning ($F = 1.47$, $df = 1/228$). Therefore, it is not surprising that incentive for the intentional items has no effect upon concurrent incidental learning ($F = .69$, $df = 1/228$). It may be concluded that incentive, as here manipulated, is not an effective variable for either intentional or incidental learning in the Type II verbal learning situation. For the purpose of evaluating other experimental variables, the normal and augmented incentive conditions may be regarded as replications of the experimental design.

*Degree of practice.*—Both intentional and Type II incidental learning increase significantly as a function of number of presentations ($F = 86.83$ and $F = 32.23$, respectively, $df = 1/228$, $P < .001$ in both cases). These data confirm earlier findings for the Type II situation (Mechanic, 1962). There are no interactions between practice and the other experimental variables.

*Orienting tasks.*—Table 1 clearly shows the predicted differences in incidental learning as a function of orienting task. With two incentive levels and two practice levels, there were four replications of the comparison among orienting tasks. In every case, incidental learning increased directly as a function of the hypothesized number of pronouncing responses. There were no reversals of order and the differences within each replication were of substantial magnitude. The overall analysis indicates the highly significant effect of orienting task on incidental learning ($F = 40.05$, $df = 2/228$, $P < .001$).

It could be argued that the effects of orienting task were not due to the hypothetical pronouncing dimension which dictated the choice of the three tasks. Rather, the tasks may have differed in time required to perform them, with concomitant variation in the amount of exposure time available for learning. If this is the case, intentional learning should show the same relation to orienting task as does incidental learning. Previous research indicates that amount of intentional learning may vary with the nature of the orienting task (Postman & Adams, 1956). Again, Table 1 shows four replications of the comparison among tasks. For three of the replications, the order for intentional learning was directly opposite to that for incidental learning. The LC $S$s showed the greatest learning while the PS $S$s showed the least learning. The fourth replication (augmented incentive-five presentations) showed a reversal with the PS $S$s scoring slightly higher than the ESP $S$s. In spite of this minor reversal, the reduction in intentional learning from LC to ESP to PS is highly significant ($F = 13.55$, $df = 2/228$, $P < .001$). This would seem to indicate that the predicted differences in incidental learning are not due to the differences in time and effort required by the various orienting tasks. If anything, predicted differences in incidental learning may be minimized because the orienting tasks requiring more pronouncing responses happen (when gauged by the intentional scores) to be more difficult, or time consuming to perform, than the tasks requiring less pronouncing responses.

It cannot be concluded that these data favor a pronouncing response hypothesis without considering still another alternative interpretation. Perhaps the effects of orienting task are a function of how much time the orienting task requires $S$ to devote to the incidental items. If an orienting task requires $S$ to spend more time on the incidental items, it is reasonable to assume that less exposure time will be available for the intentional items. This "time" interpretation can explain the finding that from LC to ESP to PS, incidental learning increases while intentional learning decreases. The data are accounted

for without utilization of the hypothetical pronouncing responses which originally dictated the choice of the three orienting tasks. However, the data in Table 1 allow a check on this interpretation by means of a further analysis. It may be noted that at five presentations, the intentional means for ESP (8.35 and 8.90) were approximately equal to the corresponding means for PS (8.15 and 9.20). Whereas intentional learning did not decrease from ESP to PS, incidental learning increased greatly (2.65 and 2.35 vs. 4.50 and 4.45). Even where the intentional scores were not lower, and could not therefore be said to reflect greater time available for incidental learning, incidental learning increased as a function of hypothesized number of pronouncing responses.

At two presentations, it appears that the "time" interpretation is consistent with the data. This interpretation is based upon the distribution of the total exposure time between intentional and incidental learning. It implies that the total of intentional and incidental learning combined should never be greater for a slowly performed orienting task than for a more quickly performed task. It was noted earlier that Task LC was more quickly performed than Task ESP which was more quickly performed than Task PS. Certainly, the time interpretation could not predict that the PS $Ss$ would learn more total items (intentional and incidental combined) than the ESP $Ss$. The ESP $Ss$ have more time for learning after performing the orienting task. Even if the ESP $Ss$, by spending less time on the orienting task, are learning less incidentally, they should at least make up for this by learning more intentional items in the greater time period remaining. The time interpretation would require that the ESP $Ss$ learn no fewer items than the PS $Ss$. The pronouncing response interpretation has no such requirement.

In order to evaluate the above argument statistically, the different incentive groups in Table 1 were combined to give 40 $Ss$ for each orienting task at each level of practice. For each $S$, the intentional

and incidental scores were combined to give a total learning score. The mean total learning for each orienting task after two presentations is 8.15, 8.50, and 8.48 for LC, ESP, and PS, respectively. Comparable means with five presentations are 12.08, 11.12, and 13.15, respectively. With two presentations, the total learning scores for the three orienting tasks do not differ ($F = .20$, $df = 2/117$). As noted above, the data for two presentations do not distinguish between the pronouncing and time interpretations. While the pronouncing view predicts differences in incidental learning, equality of total learning could still result from differences in amounts of intentional learning permitted by the different orienting tasks.

The data for five presentations argue against a time interpretation. The total learning scores of the three task groups show a significant degree of variation ($F = 4.38$, $df = 2/117$, $.01 < P < .05$). The gap between Groups ESP and PS is significant beyond the .01 level. Thus, the differences in incidental learning among the three orienting tasks are not compensated for by reciprocal differences in intentional learning. Clearly, the differences obtained as a function of orienting task are not merely due to different distributions of $S$'s time between the incidental and intentional items. It would seem more reasonable to conclude that the differences in incidental learning among the tasks are a function of the activities required by the different tasks.

*Item analysis.*—Underlying our view of the incidental verbal-learning situation is the assumption that learning—both intentional and incidental—takes place as a result of pronouncing responses made by $S$ to the stimulus items. If these same kinds of responses underlie both intentional and incidental learning, the stimuli facilitating these responses should be the same for both forms of learning. Those stimulus items most frequently learned intentionally should also show most frequent incidental learning. There are data for 24 trigrams for both intentional and incidental learning. The frequencies

with which each item was given as a correct intentional response by 120 Ss were correlated with the frequencies with which each item was given as a correct incidental response by the other 120 Ss. The product-moment correlation across the 24 items was .81. For purposes of evaluating this correlation, r's were computed for the 24 items between: (a) intentional learning at two presentations and intentional learning at five presentations; (b) incidental learning at two presentations and incidental learning at five presentations; and (c) total learning at two presentations and total learning at five presentations. The r's were .76, .59, and .82, respectively. These three r's were computed as measures of the reliability of item difficulty. Clearly, the item by item correlation between intentional and incidental learning matches the reliability indices of item difficulty. These results are consistent with the belief that the same responses are required in both intentional and incidental learning. In this view, instructions to learn increase learning because they facilitate the performance of the appropriate responses.

## SUMMARY

Type II incidental learning refers to the situation in which S is exposed to two sets of materials, instructed to learn only one of the sets, and is later tested for the materials which he was not instructed to learn. It may be expected that such learning will vary with: (a) the orienting task by which S is exposed to the incidental items; (b) the incentive for the concurrent intentional items; and (c) the number of presentations. Twelve verbal learning groups of 20 Ss each were run in a standard 3 X 2 X 2 factorial design, with the variables being orienting task, incentive, and number of presentations, respectively. Separate scores were obtained for intentional and incidental learning from each S. Incentive was found to have no effect on either intentional or incidental verbal learning. In agreement with previous results, both kinds of learning increased reliably with number of presentations. Incidental learning increased reliably as a function of the hypothesized number of pronouncing responses required by the orienting task. After considering alternative explanations, it was concluded that the nature of the responses required of S by the orienting task are of crucial importance for the amount of incidental learning which occurs.

## REFERENCES

BAHRICK, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, 47, 170–172.

BAHRICK, H. P., FITTS, P. M., & RANKIN, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, 44, 400–406.

MECHANIC, A. The distribution of recalled items in simultaneous intentional and incidental learning. *J. exp. Psychol.*, 1962, 63, 593–600.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. exp. Psychol.*, 1956, 51, 329–333.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

# RETROACTIVE INHIBITION OF R–S ASSOCIATIONS [1]

GEOFFREY KEPPEL AND BENTON J. UNDERWOOD

*Northwestern University*

Consider the S-R association that is formed during the learning of a paired-associate list (A-B). If a second list, bearing an A-C relationship to the first list (i.e., stimuli same, responses different), is interpolated between A-B learning and subsequent A-B recall, the A-B association appears to go through an extinction process (Barnes & Underwood, 1959; Briggs, 1954). This finding can lead to the assumption that extinction will occur whenever an A-B, A-C interference paradigm exists. Furthermore, negative transfer on the A-C list appears to be associated with, and may be caused by, extinction of the first-list S-R associations. Finally, it seems apparent that retroactive inhibition (RI) in recall which occurs immediately after interpolated learning, will be primarily determined by the amount of extinction occurring during the learning of the second list.

It is an accepted fact that in conjunction with the formation of S-R associations, R-S or backward associations may also develop (e.g., Richardson, 1960). In transfer studies, conflicting backward associations may be responsible for negative transfer in certain paradigms, particularly in the A-B, C-B paradigm (stimuli different, responses identical). It can be seen that if a backward association develops in learning A-B, and if, similarly, a backward association is to develop in learning C-B, the backward association of List 1 (B-A) will interfere with

the learning of the backward association in List 2 (B-C). In effect, the backward associations in the A-B, C-B paradigm form an A-B, A-C paradigm. If the R-S associations between two lists form an A-B, A-C paradigm, it seems quite possible that the R-S (B-A) associations of List 1 may also be extinguished in the same fashion as the S-R (A-B) associations. Barnes (1960) has inferred this extinction and she accounts for certain transfer phenomena by it. More particularly, Barnes inferred three different types or classes of associations which are subject to extinction whenever they complete an A-B, A-C paradigm between two lists: (*a*) forward, (*b*) backward, and (*c*) contextual associations. The present study provides a more direct test of the effect of extinction of R-S associations; in effect, it will be a study of RI of backward associations. Four paradigms will be used; if List 1 learning is symbolized as A-B, then the four paradigms may be distinguished in terms of List 2 symbols and their relation to List 1. Expectations concerning the RI of List 1 R-S associations for the four paradigms may now be stated.

*A-B, A-C paradigm.*—In learning the A-C list, no A-B, A-C paradigm is set up for the backward association, B-A; hence, no extinction of B-A is to be expected. This prediction implies complete independence between the forward and backward associations. However, to the extent that these two associations are nonindependent, RI of the B-A association will be affected by the extinction of the A-B association, since an A-B, A-C relation is obviously present for the forward association. The A-C paradigm offers an assessment of this independence since the only possible

source of extinction of the B-A association lies in the extinction of the A-B association.

*A-B, A-Br paradigm.*—In this paradigm List 2 contains the same stimuli and responses as List 1, but the stimulus and response terms are paired differently than in List 1. In learning the A-Br list an extinctive relationship (A-B, A-C) is formed for the backward association, i.e., B-A, B-Ar, so that RI of the B-A association is to be expected in this paradigm. In addition, further RI will be produced to the extent that the A-B, A-C relationship for the forward association directly affects the backward association. Since both the A-C and A-Br paradigms contain an extinctive relationship for the forward association while a similar relationship for the backward association is present in the A-Br paradigm only, a comparison between these two paradigms will provide a direct test of the hypothesis that extinction of the backward association can occur.

*A-B, C-D paradigm.*—This paradigm specifies the learning of two unrelated lists and may provide for a third type of association which is subject to extinction. Although an A-B, A-C relationship exists neither for the backward association nor the forward association for item pairs, there is the possibility of an extinctive relationship being set up between contextual cues and the stimuli of the two lists. In the learning of a list of paired associates, the responses must first be acquired before they enter into an associative connection (Underwood & Schulz, 1960). In the same manner, when the backward association is measured, the stimulus term must necessarily have been acquired as a response if it is evoked. Although it is not known to what stimulus either the response term or the stimulus term are attached when learned in this sense, it can be assumed to be some aspect of the experimental context, e.g., the memory drum, the experimental room, etc. If the association between contextual cues and the stimuli of List 1 are represented as A-B, an A-B, A-C relationship is formed during the learning of List 2 since a different set of stimuli (C) must now be attached to the same contextual cues (A). It is in this manner that the contextual association is thought to provide for a potential class of extinction-susceptible associations. At any rate significant RI of the backward association for this paradigm can be taken as support for the existence of this third class of associations.

*A-B, C-B paradigm.*—In this paradigm two types of associations which can form an A-B, A-C association may be identified and predicted to contribute to the RI of the B-A

association, namely, the backward association and the contextual association. The potential A-B, A-C relationships for these two classes of association already have been elaborated above. As with the first two paradigms listed, the C-D and C-B paradigms allow for a direct test of the extinction of the backward association; that is, the contextual association as a source of extinction is held in common for the two paradigms so that greater RI in the C-B paradigm will indicate direct extinction of the R-S association.

## METHOD

*Design.*—The design was essentially identical to that employed by Barnes (1960). Since an RI paradigm was employed, all *S*s received the same List 1; the groups were differentiated only by the construction of List 2. The four transfer paradigms of A-B, C-D; A-B, C-B; A-B, A-C; and A-B, A-Br were represented. A control group (Group C) received the single A-B list and was used to measure the amount of stimulus recall possible following an interval equal to the time that List 2 was presented to the transfer groups.

*Lists.*—The same materials constructed by Barnes were used in the present study. Each list consisted of eight nonsense syllable-adjective pairs. The stimuli for all lists consisted of nonsense syllables of 53% to 73% Glaze association value with intralist similarity and, where applicable, interlist similarity at a minimum. Responses were common adjectives having minimal intralist and interlist meaningful similarity. Four orders of presentation were constructed for each list to minimize serial learning of the responses; each order was used as a starting order an equal number of times for each condition.

*Subjects.*—Twenty *S*s were placed in each of the five groups. To assure random assignment of *S*s, the five experimental conditions were randomized 20 times so that in each of the 20 blocks of 5 *S*s, each condition occurred once. The *S*s were assigned to the experimental conditions in the order of appearance in the laboratory. The *S*s were students of undergraduate psychology courses for whom serving in experiments was a class requirement. Most *S*s had previously served in other paired-associate experiments; however, none had experienced an RI paradigm nor had been given a procedure involving stimulus recall.

*Procedure.*—The paired associates were presented on a memory drum by the anticipation method at a 2:2 sec. rate. A 4-sec.

TABLE 1

MEAN R-S RECALL

| Groups | Mean | $\sigma_m$ |
|--------|------|------------|
| A-Br | 3.40 | .34 |
| A-C | 5.05 | .43 |
| C-B | 1.85 | .42 |
| C-D | 4.60 | .45 |
| Control | 5.80 | .44 |

intertrial interval was employed. List 1 learning was taken to a criterion of one perfect recitation. For all groups except the control, a 1-min. rest was given during which time List 2 was positioned on the memory drum. List 2 was administered for 15 anticipation trials. Group C, after reaching List 1 criterion, was given a corresponding 1-min. rest and then instructions for the interpolated "rest" activity (the pyramid puzzle). The pyramid puzzle was continued for $9\frac{1}{2}$ min., an interval equivalent to the time spent by the transfer groups on List 2. Following the completion of either List 2 learning or the interpolated rest activity, a 2-min. free-recall test was presented in which S wrote down the List 1 stimuli opposite the appropriate List 1 responses listed on a dittoed sheet. The ordering of the responses on the sheet corresponded to the ordering that S would have received had List 1 learning been continued beyond criterion. An attempt was made to insure that all Ss understood what was required on the recall test prior to the commencement of the test. No S failed to reach List 1 criterion, while 1 S was dropped for failure to follow anticipation instructions, and 1 S failed to follow recall-test instructions.

RESULTS

The mean number of trials to reach criterion on List 1 for the five groups ranged from 10.55 to 15.10. An analysis of these means, however, failed to reach significance ($F = 1.68$, $df = 4/95$, $P > .05$). Although not of primary interest in the present experiment, the performance of the transfer groups on List 2 revealed Group C-D to be superior to all groups throughout the 15 acquisition trials of List 2, and Group A-Br to be generally inferior. These results are quite in line with those of previous investi-

gators (Barnes, 1960; Twedt & Underwood, 1959).

In measuring for the presence of backward or R-S associations, S was given the response terms of the A-B list and was asked to supply the stimulus terms paired with them during original learning. Only if a stimulus term was placed with its appropriate response term was a backward association said to have been demonstrated. The mean number of R-S associations for each group is shown in Table 1. The score for Group C, which is approximately 72.5% of the maximum possible (8.00), represents R-S recall without any apparent possibility of extinction of R-S associations. Inspection of Table 1 shows all transfer groups to have suffered some degree of RI when compared to Group C. An analysis of variance of the recall scores produced highly significant variation ($F = 13.70$, $df = 4/95$, $P < .001$). Comparison of means were made using the within mean sum of squares to estimate the standard error ($\sigma_{\text{diff}} = .59$); only the differences between Group A-C and Groups C-D and C did not prove to be reliable.

A problem of method arises in the interpretation of certain of the differences between means. An unavoidable procedural difference between Groups A-Br and A-C and the other three groups is the fact that the former groups received 15 more exposures to the A stimulus terms than did the latter groups. Increased exposure to the A terms during List 2 learning may make these terms more available to S at recall. This does not mean that the B-A associations will necessarily be stronger, but only that the A terms might be more available as responses at recall. This, in turn, might increase the number of "hits" if S guessed. There is, thus, the distinct possibility that, relative to

Groups C-B and C-D, the amount of RI for Groups A-Br and A-C is underestimated.

The results for the four transfer paradigms may now be assessed in terms of the three classes of associations thought to be susceptible to extinction whenever an A-B, A-C association is formed. The fact that significant RI was not obtained in Group A-C ($t = 1.27, df = 95, P > .05$) suggests that extinction of the A-B association does not result in the extinction of the corresponding B-A association as indexed by subsequent R-S recall. It is possible, however, that the methodological problem discussed above would result in an overestimation of the amount of recall for Group A-C through an increase in the number of stimulus terms available to $S$. Without any data concerning the extent to which the recall scores are inflated by guessing, it can only be tentatively concluded that extinction of the S-R and R-S associations represent two independent processes.

A comparison of the recall for Groups A-Br and A-C will not be biased due to differential stimulus-term exposure because both groups had the same number of exposures. Since both groups contain the S-R extinction factor in common, the significant difference in R-S recall for these two groups ($t = 2.80, df = 95, P < .01$) may be taken as one indication of an extinction of R-S associations when these associations form an A-B, A-C paradigm.

The final critical comparisons are between Groups C-D and C-B. Since Group C-D contains only the contextual extinction factor and Group C-B both the contextual and R-S extinction factors, the greater RI for Group C-B ($t = 4.66, df = 95, P < .01$) indicates again that the R-S association can be extinguished. The fact

that the contextual associations are involved in the recall of Groups C-B and C-D is shown by the comparison between Groups C-D and C ($t = 2.03, df = 95, .01 < P < .05$).

## DISCUSSION

This study was designed to study RI of backward (R-S) associations. The evidence presented makes it clear that when R-S associations between two successively learned lists form an A-B, A-C paradigm, RI of List 1 R-S associations may occur. Indeed, in the A-B, C-B paradigm, the magnitude of the RI was great; a difference of about four items was apparent between the control condition and recall of List 1 R-S associations with this paradigm. In line with certain previous interpretations of RI of S-R associations (Barnes & Underwood, 1959), the RI of R-S associations in the present experiment has been interpreted as being due to an extinction or unlearning of the R-S associations of the first list during the learning of the second. Certain issues related to such an interpretation require further discussion.

The fact that the A-Br paradigm resulted in a significant decrement in recall of R-S associations strongly implies an extinction process of specific R-S associations in the first list. Indeed, the amount of decrement observed is probably underestimated in the present experiment. As noted earlier, stimulus terms should be more available in this paradigm than in the control condition because of the greater number of trials they were presented. This could result in more correct R-S pairings based on guessing alone. That guessing might have occurred is suggested by the fact that in 22 instances $S$s gave a correct stimulus term but paired it with the wrong response term, and in only 1 of these 22 cases was the pairing appropriate to the association learned in List 2. Of course, if guessing did occur it may have been heightened because R-S associations of the first list had indeed been extinguished; $S$s knew the stimulus terms but extinction of the specific R-S associations

left them with no appropriate pairing tendencies. Such an interpretation might be supported by the fact that in the A-C paradigm only 9 stimulus terms were paired incorrectly and in this paradigm frequency of stimulus-term exposure was equal to that of the A-Br paradigm, but no extinction of R-S associations should have occurred.

In addition to extinction of R-S associations between pairs of items in the two lists, it has been suggested that another source of extinction may contribute to the observed recall decrement, namely, extinction of contextual associations. The extinction of such associations is said to be the sole source of decrement in the C-D paradigm. For this paradigm it is said that in learning A-B, the stimulus term A is originally associated with contextual cues, but that during C-D learning the associations between the A terms and the contextual cues are extinguished and are replaced by associations between the C terms and contextual cues. It should be noted that this process alone will not account for the decrement in recall of the B-A association for this association per se has not been extinguished. It is clear, therefore, that further assumptions are needed in order to mediate the findings if extinction of contextual associations is to be retained as a basic notion. There are several alternative assumptions which may be added and which would satisfactorily mediate the findings. However, it seems most apparent that the critical need at the present time is not further assumptions; rather, what is needed is experimental simulation of contextual cues (hence associations) in order to bring them under laboratory control and thus determine whether or not the basic notion (extinction of contextual associations) is sound.

Finally, it should be pointed out that just as in the case of presumed extinction of S-R associations (Barnes & Underwood, 1959), the reason why all R-S associations are not extinguished is not known.

## SUMMARY

This experiment investigated RI of backward or R-S associations. Various interlist relationships were studied on the premise that first-list R-S associations, like S-R associations, would be subject to extinction if the R-S associations for the two lists formed an A-B, A-C interference paradigm. All $S$s learned the same first list of paired associates which had nonsense syllables as stimulus terms and adjectives as response terms. Four different groups (20 $S$s each) then learned a second syllable-adjective list, the groups being differentiated by the nature of the interlist paradigm formed, these being: C-D, C-B, A-C, and A-Br. Following acquisition of the second list, $S$ was allowed 2 min. to write down the stimulus terms from the first list opposite the appropriate first-list response terms. A control group was given no second list.

All groups showed significant RI in R-S recall except the group learning the lists forming the A-B, A-C paradigm. No source of extinction of R-S associations for this paradigm is apparent and extinction of the S-R association does not affect the R-S association. Maximum RI was shown for A-B, C-B; this paradigm may involve extinction of specific R-S associations as well as contextual associations. The A-Br paradigm, involving only extinction of specific R-S associations, produced more RI than did the A-B, C-D paradigm; this latter paradigm is assumed to involve a decrement resulting only from extinction of contextual associations.

## REFERENCES

BARNES, J. M. "Fate" revisited. Unpublished doctoral dissertation, Northwestern University, 1960.

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, 58, 97–105.

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, 47, 285–293.

RICHARDSON, J. Comparison of S-R and R-S learning of paired-associates. *Psychol. Rep.*, 1960, 7, 225–228.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, 58, 111–116.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

# CUE SELECTION IN PAIRED-ASSOCIATE LEARNING

BENTON J. UNDERWOOD, MARGARET HAM, AND BRUCE EKSTRAND

*Northwestern University*

Consider a paired-associate list in which the stimulus term for each response consists of two distinct components, A and B. Both components are consistently present on each learning trial. Assuming that learning occurs in this situation, there are many possible interpretations which may be given as to what the effective stimulus or cue is for each response. It might be said that the effective cue is a configuration formed by A and B. Or, it might be said that each component is independently a cue for the response; that one or the other components is the effective cue, but not both, and so on.

The present study is predicated on the notion that when a complex stimulus is presented to S a selection process may occur so that the effective cue for the response is some component of the complex stimulus that is actually presented. Thus, the assumption is that there may be a discrepancy between the nominal stimulus (the stimulus actually presented S) and the functional stimulus (the component of the nominal stimulus which becomes the effective cue for response elicitation). That such discrepancies may exist is suggested by the reports of Ss that they have used only a single letter of a three-letter stimulus as the effective cue (Underwood & Schulz, 1960). Such discrepancies might also be inferred from the so-called context experiments (e.g., Weiss & Margolius, 1954) in which the removal of a component of a compound nominal stimulus produces a decrement in recall, although such studies offer other interpretative possibilities, e.g., the functional stimulus is a configuration and the removal of any component reduces the effective associative strength.

If cue selection occurs—if only a part of a compound stimulus becomes the functional stimulus—certain variables should influence the selection. The hypothesis tested in the present experiment is that given two components of different classes as the nominal stimulus, the more meaningful component will become the functional stimulus. This hypothesis seems very close to the notion of differences in discriminability as a variable determining stimulus selection, a notion suggested by Sundland and Wickens (1962), and for which some experimental support was obtained.

The particular predictions for the present study may now be specified. Two lists for original learning were constructed. The stimulus compound for one list consisted of colors and low-meaningful trigrams; for the other list the compound consisted of colors and common three-letter words. For the first list it was assumed that the colors were more meaningful than the trigrams; therefore, the functional stimuli should be the colors. In the case of the word-color compound it was assumed that the words were more meaningful than the colors, hence, the functional stimuli should be the words. (It would have been more precise to have used two sets of verbal units of known meaningfulness for the compounds, but the use of colors was recommended by the desire to keep the experiment continuous with the context experiments.) Given the

above assumptions, it was predicted that following the learning of the list with the trigram-color compounds, very little decrement would be observed if the trigrams were removed on a transfer test, but that a great loss would appear if the colors were removed from the compound. Contrariwise, in the case of the word-color compounds, removal of the words on a transfer test would result in a large loss but removal of the colors would have little effect on transfer performance.

## METHOD

The general procedure required that half the Ss learn an original list with trigram-color compound stimuli, and half learn a list with word-color stimuli. To test for each component separately on a transfer list required four groups. However, to determine precisely the amount of decrement occurring required two control groups for which the stimuli on the transfer test were the same as on original learning. Thus, the design called for six groups. Let W stand for words, T for trigrams, and C for colors, and the symbols before a hyphen designate the stimulus during original learning, those after the hyphen the stimulus on the transfer test. The six groups are, therefore: WC-WC and TC-TC (the two control groups); WC-C and TC-C (only color stimuli on transfer test); WC-W and TC-T (the verbal units appear alone as the stimuli on the transfer test).

*Lists.*—The materials for the lists are shown in Table 1. Each list consisted of seven pairs. The three-letter words all have AA ratings in the Thorndike-Lorge (1944) list except the word GAS, which is rated A. The trigrams have quite low associative con-

### TABLE 1
STIMULUS COMPONENTS USED IN THE LISTS

| Words | Trigrams | Colors |
|---|---|---|
| GAS | GWS | |
| DAY | DWK | Red |
| NEW | NXQ | Brown |
| DIE | DHX | Yellow |
| BAD | BWD | Blue |
| GOT | GVS | Orange |
| BED | BXD | Black |
| | | Green |

### TABLE 2
MEAN NUMBER OF TRIALS TO CRITERION ON ORIGINAL LEARNING AND MEAN NUMBER OF ITEMS LOST ON FIRST TRANSFER TRIAL

| Cond. | Original Learning | | Transfer | |
|---|---|---|---|---|
| | Mean | $\sigma_m$ | Mean | $\sigma_m$ |
| WC-WC | 8.80 | .97 | — | — |
| WC-W | 8.00 | 1.18 | 1.20 | .44 |
| WC-C | 9.20 | 1.14 | 2.50 | .47 |
| TC-TC | 11.55 | .95 | — | — |
| TC-T | 10.00 | 1.20 | 2.85 | .36 |
| TC-C | 10.00 | 1.21 | .05 | .37 |

nections between letters as based on the Underwood-Schulz (1960) tables. It should also be noted that the trigrams have relatively high formal similarity as indexed by repeated letters. The purpose of this was to minimize the possibility that a single letter (such as the first letter) might become the functional stimulus. The frequency of repeated letters in the word list is about the same as for the trigrams and the repetitions are in the same positions. Both lists have the same initial letters.

The color components were made of construction paper and pasted on the vellum tape. Rectangular frames of color completely surrounded the verbal unit, the width of the frame being approximately ¼ in. When the color was the only component on the transfer test the frame appeared exactly as it had during original learning, the verbal unit having been removed. The particular colors paired with particular verbal units appear in the same row in Table 1. The response terms were the single digit numbers 2 through 8. On both learning and transfer trials the rate of presentation was 2 sec. for the stimulus alone, and 2 sec. for the stimulus and response together. Four different orders of the pairs were used. Anticipation learning was used throughout.

*Procedure.*—The original learning was carried until the S achieved one perfect recitation of the list. The transfer tests were carried for 10 trials with S instructed to give as many correct responses as possible on the first transfer trial. The usual paired-associate instructions were given prior to original learning. In addition, S was told that both the word (or trigram) and color would appear consistently together from trial to trial. The intent of these instructions was to inform S of the nature of the stimulus compound with-
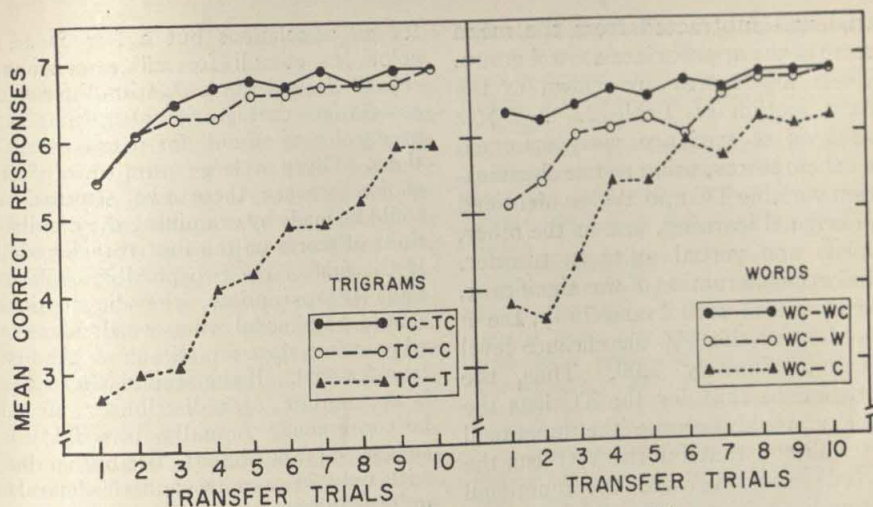
FIG. 1. Acquisition curves on the 10 transfer trials.

out, at the same time, biasing him toward "using" one or the other components. Prior to the transfer trials S again was fully informed as to the nature of the stimulus which would be present on these trials. Approximately 45 sec. elapsed between original learning and the first transfer trial.

Each of the six groups contained 20 Ss. Twenty blocks were made up such that each condition occurred once within each block, with the order of the six conditions within a block being randomly determined. The Ss were then assigned to the schedule in terms of their appearance at the laboratory. No S was lost for failure to learn.

RESULTS

The mean numbers of trials to attain the criterion on original learning are shown in the left portion of Table 2. Differences among the three WC groups and among the three TC groups represent random variation, the F being less than 1 in each case. For the 60 WC Ss the mean is 8.67 ± .63, and for the 60 TC Ss, 10.52 ± .65. The difference (1.85 ± .90) gives a t of 2.06, which is just past the 5% significance level.

The mean performance on each transfer trial is shown in Fig. 1. The left-hand section refers to the TC groups, the right-hand section to the WC groups. For the TC groups it can be seen that when the colors alone were used as stimuli, transfer was virtually complete; the performance of this group is only slightly below the control (TC-TC), thus indicating that the colors alone were completely effective functional stimuli. If *only* the color component was the functional stimulus, performance of Group TC-T should start at zero. It does not. For at least some Ss the trigrams were also functional stimuli for at least a few responses.

The right-hand section of Fig. 1 shows that for the WC groups neither the words nor the colors developed complete effectiveness. The colors are less effective than the words, but the words alone show some loss as compared with the control.

The clearest inferences concerning the functional stimuli in original learning can be made from the performance on the first transfer trial. Since Group TC-TC showed a larger criterion drop than did Group WC-WC, loss scores have been calculated. To do this, each S's score on the first

trial was subtracted from the mean score of the appropriate control group. These loss scores are shown in the right section of Table 2. A 2 × 2 analysis of variance was performed on these scores, using as one classification variable TC and WC as identified in original learning, and as the other, colors and verbal units on transfer. Only the interaction $F$ was significant, being 25.31; with 1 and 76 $df$, the $F$ needed for the 1% significance level is approximately 7.00. Thus, the predictions that for the TC lists the colors would become the functional stimuli and that for the WC lists the words would become the functional stimuli, are given some support.

## DISCUSSION

The transfer tests for the TC compounds showed that color alone was a completely effective functional stimulus. This fact precludes an interpretation of the functional stimulus as being a configuration. However, it was noted that the trigram stimuli were not completely ineffective on the first transfer trial. There are at least two possible interpretations of this finding. First, it may mean that some trigrams, quite independently of the color component, become associated directly with the response term. Secondly, it may mean that some associative connection may have developed between the color and trigram components of the stimulus compounds. Thus, when the trigram is presented, the correct anticipation is mediated by associations running from the trigram to the color to the response term. The present data do not allow a choice between these two alternative interpretations.

The transfer data following learning of the WC lists raise three interpretative problems. First, it was noted that transfer was greater when the words became stimuli than when the colors became stimuli. Two circumstances could lead to this finding. (a) Most Ss in original learning used words as stimuli

for *all* associations but a few Ss used colors as stimuli for all associations. (b) All Ss used words as stimuli for *most* associations during original learning but used color as stimuli for a few associations. Given a large number of Ss a choice between these two alternatives could be made by examining the distributions of scores on the first transfer trial. If the first of the two possible explanations is appropriate, the distributions should be bimodal when words alone or when colors alone are stimuli on the first transfer trial. If the second alternative is appropriate, each distribution should be continuous. Actually, bimodality is suggested in the present distributions but with only 20 cases in each this may be quite fortuitous.

The second interpretative problem is the same as that posed for the TC compounds where the data show that for some associations for some Ss both the trigram and the color elicited the response. Such dual functionality may also be deduced from the data for the WC compounds. On the first transfer trial a mean of approximately 5.0 correct responses occurred when the words were presented alone and 3.8 when the colors were presented alone. These two values sum to 8.8, which is appreciably higher than the mean of 6.3 shown by the control Ss on the first transfer trial. Clearly, dual functionality of the two components obtained for at least some Ss. However, just as in the case of the TC compounds, this apparent duality may result from direct associations between the components and the response term or it may result from mediation between the stimulus components.

The third interpretative problem presented by the results of the WC compounds is the fact that the words showed greater transfer than did the colors. As stated in the procedure section, it was believed that the words would be more meaningful than the colors. We have no independent evidence for this and it may not be valid. The Ss, being much more practiced in dealing with word stimuli than with patches of color as stimuli, may be biased toward the selection of the

verbal stimuli. The data are quite in harmony with such a notion. However, a somewhat different approach may be taken to the problem. An empirical test can be made to determine which stimulus compound leads to most rapid learning when this learning is not preceded by learning in which the compound is present. To determine this three new groups of 15 $Ss$ each were run. One group learned the trigram-number pairs, a second the word-number pairs, and the third the color-number pairs. The mean total correct responses in 10 trials were $38.80 \pm 3.61$, $51.20 \pm 3.52$, and $50.40 \pm 2.19$, respectively. While the $F$ is significant far beyond the 1% level it is clear that most of the variance is produced by the trigram-number pairs. The words and colors do not differ appreciably in their effectiveness as stimuli. The small difference in favor of the words occurred primarily on the first three trials. Thus, it seems quite reasonable to conclude that when $S$ is given a compound consisting of common words and colors, he is likely to select the words as functional stimuli, not necessarily because they are more meaningful, but because he is more accustomed to dealing with such stimuli.

Finally, it may be stated that the results of the present experiment, taken in conjunction with the study by Sundland and Wickens (1962), would seem to indicate that it may be more fruitful to view the so-called context experiments as experiments investigating the variables determining cue selection.

## SUMMARY

This experiment was based on the assumption that when $S$ is presented a compound stimulus in a verbal-learning experiment, cue selection may occur. Word-color or trigram-color compound stimuli were used in learning original paired-associate lists with numbers as responses, followed by a transfer test in which one or the other components alone was presented as the stimulus. Control groups were also used, these groups being given further trials with the original compound stimuli.

The results show:

1. For the trigram-color compounds, color was a completely effective stimulus on the transfer test. The trigrams, however, also produced a small positive transfer effect. The selection of the color component as the primary functional stimulus was assumed to be due to its higher meaningfulness.

2. For the word-color compounds, transfer was higher when the words appeared alone than when the colors appeared alone. This may be due to a bias $Ss$ have toward dealing with verbal material (as compared with the color patches used) rather than to higher meaningfulness of the words.

It was concluded that experiments dealing with the effects of context changes on retention may be viewed as representing cases of cue selection.

## REFERENCES

SUNDLAND, D. M., & WICKENS, D. D. Context factors in paired-associate learning and recall. *J. exp. Psychol.*, 1962, **63**, 302–306.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WEISS, W., & MARGOLIUS, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48**, 318–322.

# CONSUMMATORY AND INSTRUMENTAL RESPONDING AS FUNCTIONS OF DEPRIVATION [1]

GEORGE COLLIER

*University of Missouri*

The present study investigates the conjoint effect of deprivation on instrumental and consummatory responses.

## METHOD

*Apparatus.*—Eight Skinner boxes delivering liquid reinforcement were used (Collier & Myers, 1961). Electronic switches sensed each lick made in the magazine cup. Bar presses (BP), reinforcements, and licks were recorded by an Esterline-Angus recorder.

*Subjects.*—The 12 Ss were naive rats, 120 days old, of the Sprague-Dawley strain (Holtzman Company) maintained on Purina lab chow and tap water.

*Procedure.*—Two groups of six rats were formed. The first (H) was fed for 1 hr. after running. The second (L) had its food removed 3 hr. before running and replaced immediately after running. From Day 14 to Day 17 Group H received 15 mg. of food per day. From Day 17 to the finish they received 10 gm. per day. These changes were required to keep Ss' weights relatively constant.

Each S received 5 days of table training, handling, and deprivation accommodation followed by 4 days of magazine training, 10 days of reinforced BP on a 1-min. fixed interval (FI), and 3 days of extinction. During reinforced bar pressing all Ss were given .3 ml. of 16% sucrose solution per reinforcement. Thus, since by the fifth day of bar pressing all reinforcements possible were presented and all those presented were taken, each S received 15.6 ml. of solution per day in the 52-min. test period.

All Ss were weighed every day immediately before running.

## RESULTS

*Reinforced bar pressing.*—The lower panel of Fig. 1 presents the weights of the two deprivation groups over the course of the experiment. Group H ranged from approximately 86% of the body weight of Group L at the beginning of the study to approximately 80% at the end. The weight of Group L increased significantly across sessions. The upper panel of Fig. 1 presents the total number of BP in 52 min. for each of the deprivation groups over the course of the experiment. It is clear that Group H was superior to Group L under all circumstances.

An analysis of variance of the number of BP and of licks on Day 17 shows that for bar pressing the main effects of Deprivation ($F = 22.6$, $df = 1/10$, $P < .01$) and of Minutes ($F = 2.43$, $df = 51/510$, $P < .01$) were significant, while for licks the main effect of Minutes only ($F = 1.46$, $df = 51/510$, $P < .05$) was significant. For neither measure did the interaction of Deprivation $\times$ Minutes
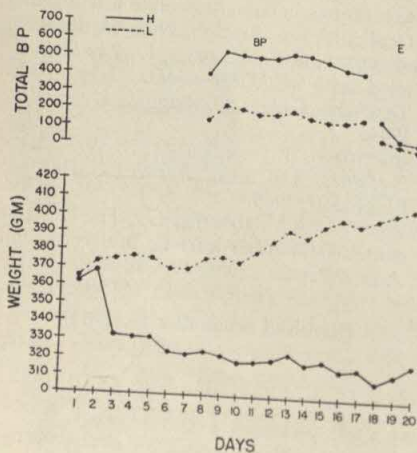


FIG. 1. Total daily bar presses and weight of Ss as a function of deprivation.
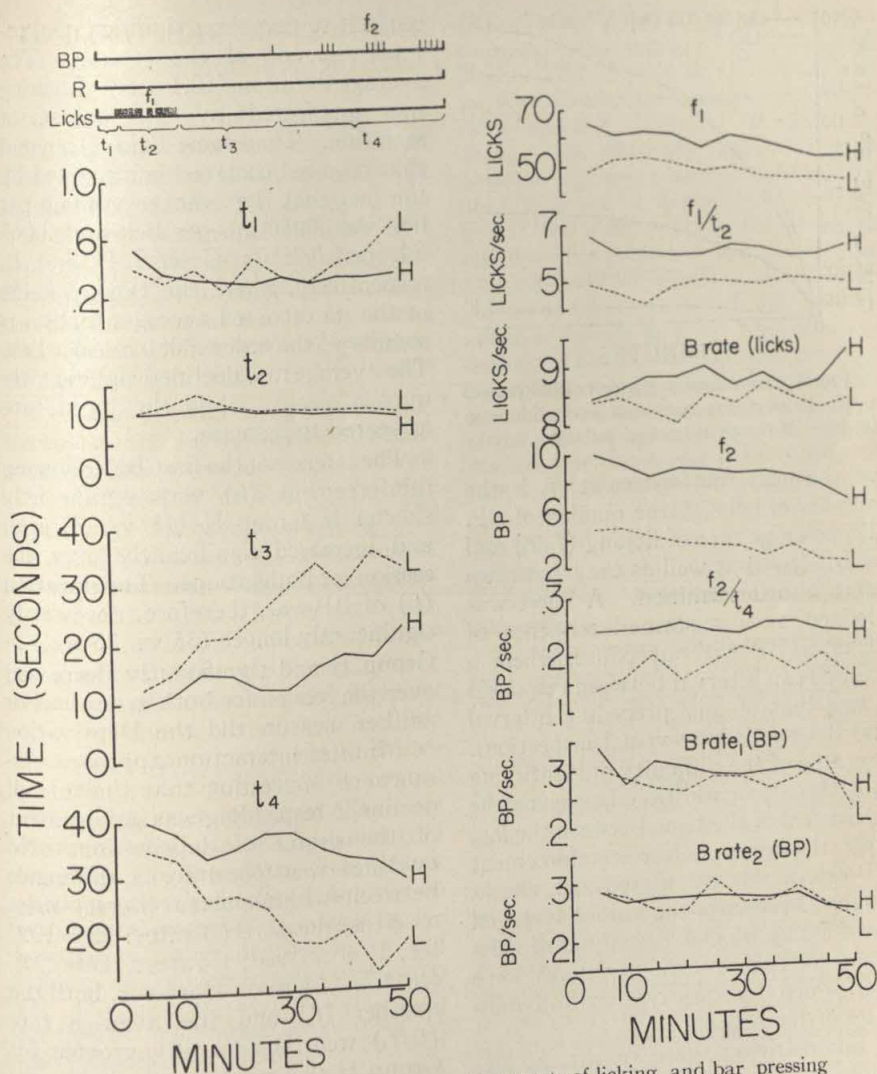
FIG. 2.  Latency, duration, rate, and burst rate of licking, and bar pressing as functions of deprivation and satiation.

approach significance. The total numbers of BP per session were 439.7 and 174.0 for Groups H and L, respectively, while the total numbers of licks were 2763 and 2400, respectively.

In order to assess the locus of the rate differences and changes observed, the pattern of responding in each minute of the FI schedule, i.e., between reinforcements, was examined. Figure 2 (top left) presents a schematic of 1 min. (between reinforcements) of a typical tape from the Esterline-Angus records: $t_1$ is the time from the reinforcement to the first lick, $t_2$ is the duration of licking, $t_3$ is the time from the last lick to the first BP, $t_4$ the time from the first BP
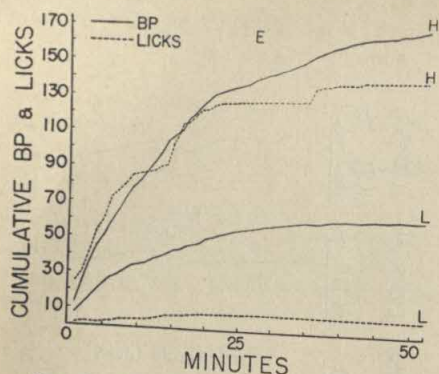
FIG. 3. Cumulative number of bar presses and licks over the course of extinction as a function of deprivation and satiation.

to the next reinforcement; $f_1$ is the number of licks, $f_2$ the number of BP. The average rate of licking $(f_1/t_2)$ and of BP $(f_2/t_4)$ as well as the burst rates were also determined. A burst was defined as a sustained sequence of licks or BP without pause, where a pause is an interval between responses twice that of any preceding interval (as determined by visual inspection). Because of the difficulty of identifying a BP burst, an additional measure, the reciprocal of the time between the last two responses before reinforcement (B-rate$_2$), was calculated as a check. Figure 2 presents the values averaged over 5-min. blocks. Analyses of variance on each of the measures were performed for deprivation and minutes of the session.

Summarizing these results, no significant effect of Deprivation on the average (H = 6.0/sec, L = 4.8/sec) or Momentary (H = 8.8/sec, L = 8.4/sec) Rate of Licking was found; however, there was a suggestion that there might be a difference in average rate as a function of a greater number or duration of pauses between bursts of licks at a constant rate ($F = 4.35$, $df = 1/10$, $P < .10$). The initial latencies of the first licks were significantly longer at the beginning and

tended to increase significantly toward the end of the session. The average duration (10.1 sec.) of licking was unaffected by the course of satiation. That some licking beyond that required occurred is indicated by the fact that the average volume per lick was .0056 ml. per lick and .0065 ml. per lick for Groups H and L, respectively, where the typical value in the unrestricted access situation is usually of the order .0080 ml. per lick. The average rate declined significantly over a session while the burst rate appeared to increase.

The latency of the first BP following reinforcement ($t_3$) was significantly shorter in Group H (15 vs. 26 sec.) and increased significantly over the session for both groups. The duration ($t_4$) of BP was, therefore, necessarily significantly longer (35 vs. 23 sec.) in Group H and significantly decreased over the session for both groups. For neither measure did the Deprivation × Minutes interaction approach significance, suggesting that the rate of decline in responding was independent of the degree of deprivation. No evidence was found for a difference between within-cluster (burst) rates on either measure (B-rate$_1$: H = .29/sec, L = .27/sec; B-rate$_2$: H = .27/sec, L = .24/sec). However, both the number ($f_2$) and the average rate ($f_2/t_4$) were significantly greater for Group H (8.4 vs. 3.5 and .24/sec vs. .17/sec, respectively). The average rate did not decline over the session while the frequency of responding did. The difference between the average rate of responding and the momentary rate reflects either the burst duration, interburst interval, or both. The decline in frequency across the session appears to be solely due to the decreasing time available for BP because of the increasing latency of the first BP.

*Extinction.*—The cumulative form of the extinction curve is given in Fig. 3. Both the total number of BP and total number of licks differed significantly as a function of deprivation. The rate of decline in bar pressing was significantly faster for Group L as was the interaction of minutes and deprivation. No analysis of the effect of time and its interaction with deprivation was performed for licks since Group L stopped licking within the first few minutes.

Two things were clear in the extinction data. First, the rate at which consummatory responding as well as bar pressing declined depended strongly on deprivation, stopping with a few licks of the empty magazine at low deprivations. Second, bar pressing fell off at a much slower rate than licking, substantial amounts of bar pressing occurring after the licking response to the magazine had stopped.

## DISCUSSION

When the distribution of responses between reinforcements is examined it is clear that the momentary rate of responding for both the consummatory and instrumental response is relatively insensitive to either deprivation or the number of reinforcements consumed. The difference in average rate as a function of deprivation is the result of differences in the duration, or interval between bursts, or both, and the differences in overall rate of BP as a function of deprivation result from both the latency of the first response following reinforcement and the differences in average rate. The decline in the overall rate of BP as a function of the number of reinforcements consumed appears to be solely a function of the latency of the first response following reinforcement and the rate of decline is independent of the degree of deprivation. This distribution of responses has been reported previously for the consummatory re-

sponse when deprivation and amount consumed are examined (e.g., Davis & Keehn, 1959; Stellar & Hill, 1952) and for the instrumental response on a fixed-ratio schedule (Sidman & Stebbins, 1954). The relatively small difference between average rates of licking, the fixed duration of licking, and the apparent lack of difference in the latency of the first lick probably result from the schedule used in which a fixed, small volume is presented which can be consumed in a single or a few brief bursts of licking. That is, the duration required is within the range of durations of bursts of licking. The failure of the latency of the first lick to vary as a function of the amount consumed (see Stellar & Hill, 1952) is probably the result of the fact that the consummatory response in the Skinner box situation is under the control of a precise discriminative stimulus, namely the operation of the magazine.

Ingestion ceases before nutrition has been accomplished. This fact suggests that satiation and deprivation are two distinct, independent processes (Collier & Myers, 1961). The results of the present study provide further evidence for this distinction. They suggest that satiation, which varies as a function of the momentary postingestive load, evinces its major effect on the latency of the response measured, while deprivation affects both the latency and burst density. The present data further suggest the possibility that the parameters of the consummatory response are insensitive to such variables as deprivation and satiation, and thus it is the ancillary food approaching and food producing responses which determine the rate, pattern, and amount of consumption.

## SUMMARY

Twelve albino rats were divided into two deprivation groups, high and low, and were run in Skinner boxes for a .3-ml., 16% sucrose reinforcement. Analysis of the grain of the consummatory (licks) and instrumental (bar press) responses led to the conclusion that for the consummatory response neither the

latency, duration, average, nor momentary rate of responding varied significantly as a function of deprivation. Latency showed a slight tendency to decrease for high deprivation and increase for low deprivation, average rate to decrease, and momentary rate to increase across a session. For the BP response, on the other hand, the latency and therefore necessarily the duration of pressing were functions of both deprivation and satiation. The average rate proved to be related to deprivation only, while two measures of the momentary rate appeared related to neither deprivation nor satiation. In extinction the rate of occurrence of licking showed a much more rapid decline than did rate of BP.

## References

Collier, G., & Myers, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61**, 57–66.

Davis, J. D., & Keehn, J. D. Magnitude of reinforcement and consummatory behavior. *Science*, 1959, **130**, 269–271.

Sidman, M., & Stebbins, W. C. Satiation effects under fixed-ratio schedules of reinforcement. *J. comp. physiol. Psychol.*, 1954, **47**, 114–116.

Stellar, E., & Hill, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

# DEPTH PERCEPTION IN ROTATING DOT PATTERNS: EFFECTS OF NUMEROSITY AND PERSPECTIVE [1]

MYRON L. BRAUNSTEIN [2]

*University of Michigan*

The changing shape of a visual pattern may elicit a three-dimensional perception, even when other cues to depth are so reduced that any stationary view of the pattern appears two-dimensional. Stimuli for the study of this effect have generally been produced by rotating an object or pattern between a light source and a translucent screen, thus displaying shadows of the transforming pattern. This method was employed by Miles (1931), who used a two-bladed electric fan as a shadow caster, and by Metzger (1934), who used a number of cylinders to generate his displays.

The effect was systematically investigated by Wallach and O'Connell (1953), who termed it the "kinetic depth effect," and by Gibson and Gibson (1957). While earlier investigators regarded the depth effects created by such displays as illusions, recent experimenters have recognized that they involve a cue to depth not fully incorporated into the classical set of depth cues. Gibson (1950) has aptly termed this cue "motion perspective."

Green (1959a, 1959b) introduced a method which allows greater variety

in the presentation of stimuli and greater control of the motion of parts of a display than the shadow projection method. By means of instructions to a high speed digital computer equipped with a CRT output recorder, a motion picture can be made of a two-dimensional projection of any mathematically specifiable transformation of any figure. Green has studied the effects of numerosity, speed, and axis of rotation on subjective judgments of the extent to which the parts of the display maintain the same relative position, i.e. coherence, for filmed sequences representing projections of points or of line segments rotating in three-dimensional space.

*Definition of perspective.*—The perspective used in producing a two-dimensional projection of a three-dimensional display may be conveniently defined as the ratio of the distance between the projection point and the most distant $X–Y$ (frontal parallel) plane to the distance between the projection point and the closest $X–Y$ plane, along the $Z$ axis (line of sight). This is equivalent to the ratio of the projection of a distance on the closest $X–Y$ plane to the projection of the same distance on the most distant $X–Y$ plane. Figure 1 illustrates these definitions of perspective for a four-point pattern confined to an imaginary cube.

If the projection point is at infinity and thus equally distant from all elements of the pattern, the perspective ratio is 1, and the projection of the pattern does not change with its location in depth. This method of
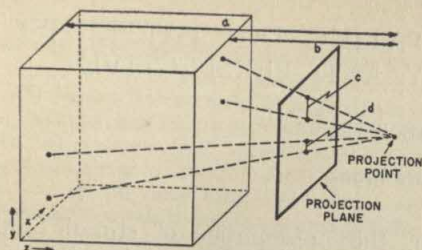
Fig. 1. Projection of points in a cube. (Perspective is defined as $a/b$, or equivalently, $c/d$.)

projection is referred to as parallel projection. If the ratio is greater than 1, the projection of the pattern varies with its location in depth, and the projection is referred to as polar.

*Motion perspective.*—Motion perspective is a complex cue to depth involving several better known aspects of perception. If a polar projection of a three-dimensional pattern rotating about an axis perpendicular to the line of sight is displayed, the rate of motion of the projection of a part of the pattern is an inverse function of its distance from the projection point. This is essentially motion parallax, in the classical sense. Also, in the case of a polar projection, the distance between the projection of two parts of a pattern is an inverse function of the distance of those parts of the pattern from the projection point. This cue is similar to linear perspective, except that it involves temporal as well as spatial variations in the display.

With both polar and parallel projections, changes in the relative distances between the projections of parts of the pattern correspond to changes in orientation of the pattern from a fixed projection point. In the case of parallel projections, this effect is not confounded with motion parallax and linear perspective. Although only very distant objects are seen in

parallel projection in direct vision, the present study compares polar and parallel projections in order to separate the more familiar depth cues from the novel aspects of motion perspective.

As the projection point is equivalent to the nodal point of the observer's eye in retinal projections, variations in perspective are normally accompanied by changes in the visual angle subtended by the pattern. In the present study the size of the projection was controlled by placing $O$'s eye at a fixed intermediate distance from the computed location of the three-dimensional pattern, and not at the computed projection points. Perspective is thus "inappropriate" to $O$'s distance from the computed location of the pattern. It is exaggerated by closer projection points and reduced by more distant ones.

## METHOD

*Stimuli.*—The stimuli were produced by Green's (1959a) computer method. Specifically, from two to six points in a cube having sides of two units, centered about the origin, were randomly selected. A projection of the points onto a specified plane perpendicular to the $Z$ axis was computed, using a projection point having a $Z$ coordinate of 2, 4, 16, or 512, and $X$ and $Y$ coordinates of 0. The projection plane was always equidistant from the projection point and the origin so that a distance $d$ on the plane $Z = 0$ was projected into a distance $d/2$, regardless of the distance of the projection point from the origin. Projections were computed at each 4.5° as the hypothetical cube was rotated 360° about the $Y$ axis. There were thus 80 projections for each display. Each projection was plotted on the face of a CRT using an overlapping cluster of four spots to represent a point. These clusters, hereafter referred to as spots, were roughly circular in shape and were approximately .01 unit in diameter. The plots were photographed on 16-mm. film and each sequence of 80 photographs became a stimulus display in the final film.

Forty such displays were prepared, representing 20 treatment combinations (5 levels of numerosity and 4 of perspective), each

produced twice, with different selections of random points. The stimulus films consisted of pairs of these displays. Each pair was composed of an 80-frame display, 28 frames of blank film of the same optical density as the background of the displays, a second 80-frame display, and 108 frames of blank film, again of the same density as the background of the displays. Presented at 16 frames per sec., this would mean a 5-sec. display, a 1.75-sec. pause, another 5-sec. display, and a 6.75-sec. pause, followed immediately by the first display in the next pair.

*Apparatus.*—The films were projected by a Kodak Royal 16-mm. projector in which a 500-w. bulb was used. The lens had a focal length of 2.5 in. The displays were projected onto a translucent plastic screen, 174⅞ in. from the front of the lens. One unit on the theoretical projection plane became 3.76 in. on the screen. The *O* viewed the displays from the opposite side of the screen. Background and spot luminances were approximately .009 and 1.1 ft-L, respectively. The projector operated at a speed of 16.2 ± .1 frames per sec. The *E* operated the projector and recorded *O*'s responses, which were transmitted to a tape recorder in *E*'s cubicle.

The *O* viewed the displays monocularly through a reduction tunnel. This consisted in part of a cylinder 4⅞ in. in length and 2½ in. in diameter, against which *O* placed his eye. The cylinder was covered on the outside with black tape and lined with black paper on the inside. It was inserted into a circular aperture 2⅝ in. in diameter in a wooden disk 15 in. in diameter and 1½ in. thick. The length of the cylinder inserted into the aperture was ¾ in., leaving 4⅛ in. protruding on the side on which *O* viewed the displays. Black cloth, formed roughly into a cylinder, covered the space between the disk and the aperture in the wall in front of *O*, which was also 15 in. in diameter. The distance from the front of the disk to the wall was 21⅞ in. The *O*'s eye was thus approximately 26 in. from the screen, or 13.8 theoretical units from the origin of the display. The *O* could see neither the borders of the aperture in the wall as he looked into the reduction tunnel, nor the location of the screen as he entered the room. The *O*s were 31 male students enrolled in the elementary psychology course; 20 served in Exp. I and 11 in Exp. II.

*Procedure.*—Each *O* was tested individually in a 2-hr. session. In Exp. I, *O* was told that he would see pairs of displays, either of which could be seen as occurring in space or on a flat surface in front of him, and was to indicate, by saying "first" or "second," which display in each pair gave the stronger impression of occurring in space. After a practice reel, *O* was shown four reels containing the 190 pairs of the present experiment, as well as 34 additional pairs not included in this report.

The procedure in Exp. II was identical to that used in Exp. I, except for a change in the critical portions of the instructions. For each pair of displays, *O* was asked to decide which display showed the greater coherence, with coherence defined as "the degree to which the parts of the display seem to maintain the same relative positions as the display moves." No mention was made of the possibility of seeing the displays either as two-dimensional or three-dimensional.

## RESULTS

Each treatment combination was represented in 19 paired comparisons. For each of the 20 *O*s, the number of times a stimulus representing each treatment combination was chosen[3] was calculated. The method employed to analyze these frequencies is a direct extension to three factors (numerosity, perspective, and *O*s) of the median test described by Mood (1950) for "two factor experiments, one observation per cell" (pp. 399–402).

In Exp. I significant effects were found for numerosity $(P < .05)$ ($\chi^2 = 125.96$, $df = 4$), perspective $(\chi^2 = 21.90$, $df = 3$), and the interaction of the two treatment dimensions $(\chi^2 = 108.30$, $df = 19$). The proportion of trials on which stimuli representing each of the treatment combinations were chosen by all *O*s is given in Table 1. A greater effect of perspective is indicated for the stimuli having smaller numbers of points. The relationship between the propor-

[3] "Chosen" is used to mean "chosen as giving a stronger impression of occurring in space than the stimulus with which it was paired" when reference is made to Exp. I, and "chosen as appearing more coherent than the stimulus with which it was paired" when reference is made to Exp. II.

MYRON L. BRAUNSTEIN

TABLE 1

PROPORTION OF TRIALS ON WHICH STIMULI WERE CHOSEN AS GIVING A STRONGER DEPTH IMPRESSION: EXP. I

| Perspective | Number of Points | | | | | All Numbers |
|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | |
| 1.00 | .232 | .329 | .445 | .524 | .561 | .418 |
| 1.13 | .245 | .508 | .547 | .537 | .632 | .494 |
| 1.67 | .358 | .376 | .621 | .634 | .589 | .516 |
| 3.00 | .447 | .524 | .589 | .608 | .695 | .573 |
| All levels | .320 | .434 | .551 | .576 | .619 | |

TABLE 2

PROPORTION OF TRIALS ON WHICH STIMULI WERE CHOSEN AS APPEARING MORE COHERENT: EXP. II

| Perspective | Number of Points | | | | | All Numbers |
|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | |
| 1.00 | .708 | .598 | .517 | .665 | .541 | .606 |
| 1.13 | .685 | .594 | .517 | .527 | .603 | .585 |
| 1.67 | .541 | .522 | .483 | .503 | .483 | .506 |
| 3.00 | .527 | .273 | .287 | .230 | .201 | .303 |
| All levels | .614 | .496 | .451 | .481 | .457 | |

tion of times stimuli representing each level of numerosity were chosen and the logarithm of the number of points for that level, is illustrated in Fig. 2. Figure 3 illustrates the corresponding relationship for levels of perspective.

The frequencies of trials on which stimuli representing each treatment combination were chosen by each of the 11 $O$s in Exp. II were also subjected to a median analysis. Significant effects ($P < .05$) were found for numerosity ($x^2 = 25.59$, $df = 4$), perspective ($x^2 = 42.27$, $df = 3$), and their interaction ($x^2 = 63.22$, $df = 19$). The proportion of trials on which stimuli representing each of the treatment combinations were chosen by all $O$s is given in Table 2. The relationships between numerosity and proportion of choice and between perspective and proportion of choice are illustrated in Fig. 2, and Fig. 3, respectively.

Figures 2 and 3 also illustrate the opposite effects of the treatment dimensions on the two dependent variables, depth judgments and coherence judgments. As well as difference in direction of effect, the treatment dimensions showed differences in relative importance in their effects on the two types of judgments. The 20 proportions in Table 1 are ordered primarily on the basis of numerosity and secondly on the basis of perspective, while the proportions in Table 2 are primarily ordered on perspective and secondly on numerosity.

DISCUSSION

The proportion of trials on which a stimulus was judged to give a stronger impression of occurring in space than the stimulus with which it was paired was found to be an approximately linear function of the logarithm of the number of points in the stimulus. This supports



FIG. 2. Effect of numerosity on depth and coherence judgments.
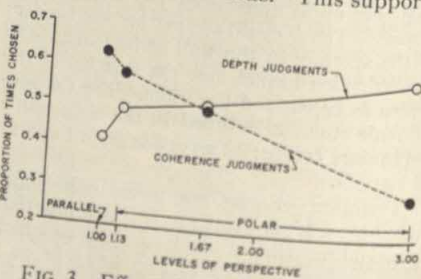


FIG. 3. Effect of perspective on depth and coherence judgments.

the hypothesis of a direct relationship between number of points and judged strength of the depth impression. Although the logarithmic relationship was not specifically predicted, it is the same as that obtained by Green (1959a) using larger numbers of spots and ratings of subjective coherence.

Perspective significantly affected the choice of stimuli giving stronger depth impressions, although it appeared to be secondary to number of points in its effect on the ordering of the stimuli. This would indicate that the interrelated factors of motion parallax and linear perspective do affect the depth impression created by these displays.

The question arises as to whether number of points may be important only because increasing numbers of points provide additional opportunities to view the effects of perspective. This is contraindicated by the findings that larger differences between levels of perspective occurred for smaller numbers of points and that the relationship between number of points and proportion of choice holds even for parallel perspective. It would appear instead that perspective becomes more important as the cues to change in orientation of the pattern are reduced, i.e., with smaller numbers of points.

Some insight into the role of subjective rigidity or coherence of the patterns in influencing perceived depth may be gained from the results of the second experiment, which showed different effects of the independent variables when coherence judgments were used as the dependent variable. Displays of two spots were judged most coherent. This would be expected, as two spots moving on a plane could always represent the projections of two points at a constant distance in three dimensions. The relationship between numerosity and coherence judgments for larger numbers of spots was not very pronounced.

A much greater effect on coherence judgments was produced by perspective. Stimuli produced with closer viewing points were judged less coherent than those with more distant viewing points.

There was thus an inverse relationship between depth judgments and coherence judgments across levels of perspective. This would seem to result from a combination of two factors. First, as all stimuli were not perceived as three-dimensional at all times, the changes in distances between the two-dimensional projections of the points may have influenced average coherence judgments, and more variability in the two-dimensional projections did occur with greater perspective. The other factor is the distortion resulting from the magnification and demagnification of the projections used to maintain a uniform "average" size. This would especially affect the stimuli with maximum perspective, which differed furthest from appropriate perspective, and the difference between depth and coherence judgments is greatest in this case.

The use of perspective has been an area of disagreement between J. J. Gibson and H. Wallach, the two major experimenters in this area. Gibson (1957) uses what he properly regards as appropriate perspective in his displays. A translucent screen is placed *equidistant* between O and a point light source and shadows of objects rotating between the light source and screen are observed. This is equivalent to displaying projections computed with the projection point at the position at which O will place his eye when he observes the stimuli. Wallach (1953), on the other hand, uses a distant projection point which is further from the screen than O's eye. The resulting situation is similar to the use of parallel perspective in the present research.

Gibson's situation is more like the real-life perceptual situation and present results would indicate that it should in general lead to a stronger impression of depth than does Wallach's method. The "kinetic depth effect," as formulated by Wallach, does however require parallel projections of the three-dimensional displays for its isolated study.

The choice of perspective should then be made according to the purpose of the experimenter in these respects. The

present results indicate that the importance of variations in perspective will depend upon the type of judgments used. When the effects of variation in perspective are of interest, at least for displays involving small numbers of elements, depth judgments and rigidity judgments cannot be assumed equivalent.

## SUMMARY

Motion picture sequences of spots representing projections of points rotating in three dimensions were produced using the CRT output of a digital computer. The sequences varied in number of points and in perspective. Perspective is defined as the ratio of the distance from the projection point to the most distant plane of the display to the distance from the projection point to the closest plane of the display. A paired comparison method was used to elicit judgments of the relative strength of the depth impressions created by the sequences and of the relative coherence of the patterns while in motion.

Judged strength of the depth impression increased with increasing numbers of spots and, to a lesser degree, with increasing perspective. Subjective coherence decreased with increasing perspective. The role of perceived coherence of moving patterns in depth perception and the effects of variations in perspective on depth judgments are discussed.

## REFERENCES

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

GREEN, B. F. Kinetic depth effect. (Psychology Group 58) Quarterly Progress Report, Massachusetts Institute of Technology, Lincoln Laboratory, 1959. (a)

GREEN, B. F. Mathematical notes on 3-D rotations, 2-D perspective transformations, and dot configurations. Group Report No. 58-5, Massachusetts Institute of Technology, Lincoln Laboratory, 1959. (b)

METZGER, W. Tiefenerscheinungen in optischen Bewegungsfeldern. *Psychol. Forsch.*, 1934, **20**, 195–260.

MILES, W. R. Movement interpretations of the silhouette of a revolving fan. *Amer. J. Psychol.*, 1931, **43**, 392–405.

MOOD, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

WALLACH, J., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205–217.

# SUPPLEMENTARY REPORT: PROACTIVE INHIBITION AS A FUNCTION OF THE METHOD OF REPRODUCTION [1]

JOHN L. WIPF AND WILSE B. WEBB

*University of Florida*

Greenberg and Underwood (1950) found that paired-associate recall, after 5 hr., was inversely related to the number of lists previously learned. Thus, proactive inhibition occurred. Their learning materials were two-syllable adjectives arranged in four paired-associate lists of 10 pairs each. The anticipation method was used to test recall.

*Method.*—The present study is a replication of the Greenberg and Underwood study in that it employed a 5-hr. retention interval, a 2-sec. presentation rate on a memory drum, the same learning-relearning criteria (8/10 correct and 10/10 correct, respectively), and four lists over 4 days. However, in the present study, the method of presentation was modified to permit the use of free recall or reproduction as a retention measure. The 10 Greenberg-Underwood response words in a list were shown one at a time. After each trial (one presentation of the 10 words), *Ss* were required to reproduce in writing the words they could remember. A 3-min. limit was set on this recall. Successive trials, using four different serial arrangements, were given until learning criterion was met.

The method of reproduction was similarly employed in testing recall, i.e., after 5 hr. *Ss* reproduced, in writing, as many words as they could recall. Recall was immediately followed by relearning. A different list was used on each of 4 successive days. Twenty-four different orders of the four lists were used. Twenty-four *Ss* (11 men and 13 women introductory psychology students) each learned the lists in a different order.

*Results.*—Separate analyses of three dependent measures, number of trials to learn, number of words correctly recalled, and number of trials to relearn, were performed. The means and *SD*s of these measures are given in Table 1. An analysis of variance showed the differences among the mean number of words correctly recalled over days to be significant (*P* < .001). Thus, significant PI at recall occurred from day to day, after the first day. Recall differences among *Ss*

### TABLE 1

MEANS AND *SD*S OF TRIALS TO LEARN, WORDS RECALLED, AND TRIALS TO RELEARN

| Day | Trials to Learn | | Words Recalled | | Trials to Relearn | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| 1 | 2.46 | .76 | 6.17 | 2.15 | 3.00 | 3.27 |
| 2 | 2.91 | 1.55 | 4.95 | 2.42 | 2.79 | 2.25 |
| 3 | 2.58 | .81 | 4.62 | 2.64 | 2.17 | 1.11 |
| 4 | 2.21 | .91 | 3.92 | 2.44 | 2.58 | 1.71 |

were significant. Recall differences among lists were not significant.

The analysis of the learning measures showed differences due to days to be significant (*P* < .05). These differences appear to result from facilitation effect (see Table 1 means, Days 2 to 4). List differences were significant (*P* < .01). Thus, word lists differed in difficulty in terms of trials-to-learn. Subject differences were again significant.

List difficulty in terms of trials-to-learn should have had no differential effect because all possible sequences of the four lists were used; hence list differences were counterbalanced throughout the study.

The analysis of the relearning measures showed Day and List differences not to be significant. Subject differences were significant.

Allowing ample time for response and written reproduction, the use of unpaired presentations increased the number of words recalled and decreased the number of trials required to learn and relearn (compare Table 1 with the Greenberg-Underwood results). However, a PI effect was demonstrated. A minimal conclusion which may be drawn from these data is that the PI effect demonstrated by Greenberg and Underwood is not a function of the particular learning-recall procedure utilized.

### REFERENCE

GREENBERG, R., & UNDERWOOD, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, 40, 452–457.

# SUPPLEMENTARY REPORT: EFFECT UPON SENSORY PRECONDITIONING OF BACKWARD, FORWARD, AND TRACE PRECONDITIONING TRAINING [1]

JAMES D. WYNNE AND W. J. BROGDEN

*University of Wisconsin*

Hoffeld, Thompson, and Brogden (1958) report variation in magnitude of sensory preconditioning (SP) in cats as a function of stimuli time relations during preconditioning. The 200 preconditioning training trials involved a 2-sec. light paired with a 1000-cycle tone either of 2, 2.5, 3, 4, or 6 sec. duration. Since tone and light terminated together, the independent variable of tone precedence is represented by values of 0, .5, 1, 2, and 4 sec. Significant linear and quadratic trend was found for magnitude of SP with the maximum at 4 sec. precedence. There is, however, a confounding of precedence and duration, and the longer durations of tone represent conditions for delayed conditioning. If a fixed duration of tone is used, there is no confounding of duration with precedence, but the large values of precedence then represent the conditions of trace conditioning.

*Method.*—The present experiment is a replication of the experiment reported by Hoffeld et al. (1958) except for the following differences: (*a*) duration of the tone was always held constant at 2 sec.; (*b*) the range of precedence of tone was increased and included negative values (backward conditioning) because Silver and Meyer (1954) report equal magnitudes of SP for simultaneous and a backward preconditioning time relation; (*c*) the number of preconditioning training trials was reduced from 200 to 4 because Hoffeld, Kendall, Thompson, and Brogden (1960) found maximum SP for the latter value; (*d*) *N* for each group was 8; (*e*) one session of adaptation to the rotator was introduced; and (*f*) following the test for SP all *Ss* were given tone-shock avoidance training in the attempt to obtain an additional measure of SP. Since this procedure produced no significant results, no report of them is made.

*Results.*—Magnitude of SP is measured by frequency of response in the cross-modal generalization tests of tone alone, given following avoidance conditioning to the light CS. An orthogonal polynomial analysis of variance (Grant, 1956) of the data for the experimental groups shows significant differences between groups and a significant quadratic trend (*P* = .05). The experimental group means are plotted in Fig. 1 as is also the curve of the best-fitting quadratic equation. Range tests (Duncan, 1951) indicate that the mean for 4 sec. precedence is significantly different from all other means, none of which differ significantly from each other. The means for precedence conditions −4, −2, and 16 sec. are either approximately equal to or less than the mean for the control group (.9), so it is improbable that any SP occurs for these conditions.

The control group of the present study shows greater cross-modal generalization than similar groups have shown in earlier studies (Brogden, 1939; Hoffeld et al., 1958). If the control *Ss* of Hoffeld et al. (1958) are added to the present control group, *N* is 12 and the mean is .58. The upper fiducial limit for cross-modal generalization provided by the *t* test is .83 at the 5% level. This limit clearly excludes evidence of SP at precedence levels of −4 and −2 sec. (backward conditioning) and long trace conditioning at 16 sec. precedence. If the combined control group data are used in one-tailed *t* tests with that of the remaining experimental groups, the means for precedence conditions 0, 1, 2, and 4 sec. exceed the control group mean at the 5% level, and for precedence conditions −1 and 8 sec.
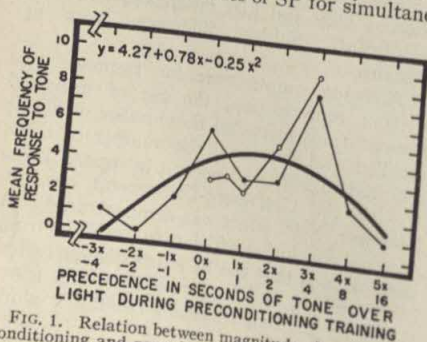


FIG. 1. Relation between magnitude of sensory preconditioning and precedence of tone over light during preconditioning. (The data represented by the open circles is taken from Hoffeld et al., 1958. The filled circles represent the data of the present experiment. The smooth curve is the best fitting quadratic function for the data of the present experiment, the equation for which is given in the upper left corner of the figure.)

exceed the control mean at the 20% and 10% levels, respectively.

*Discussion.*—The present results are in good agreement with those of Hoffeld et al. (1958) which are plotted in Fig. 1. In comparing the two studies it is evident that the confounding of duration of tone with its precedence ove light during preconditioning training has little if any effect upon the magnitude of SP. Saying it another way, delay and trace conditioning procedures during preconditioning make little if any difference as long as the time relations between the onset of the CS (tone) and onset of the UCS (light) are the same.

A difference in magnitude of SP, favoring the present experiment, was expected on the basis of parametric differences in the number of preconditioning trials, since Hoffeld et al. (1960) found this variable to have a marked effect upon the magnitude of SP. They obtained a mean frequency of response to trials of tone alone of 24.17 for 4 preconditioning trials (value used in the present study) and a mean of 10.50 responses for 200 trials (value used by Hoffeld et al., 1958). The time relations during preconditioning of the Hoffeld et al. (1960) experiment were those found optimal by Hoffeld et al. (1958) and involved a 6-sec. tone and 2-sec. light with 4 sec. precedence. In view of the minor differences in amount of SP for comparable time relations between the present study and that of Hoffeld et al. (1958), the possibility of interactions between trace and delayed time relations with number of preconditioning trials must be considered. The possibility also exists that an unknown variable or parameter present in the Hoffeld et al. (1960) study accounts for the high level of SP.

Although a significant quadratic trend in magnitude of SP as a function of time relations was found in the present experiment, the best-fitting quadratic equation does not give a satisfactory description of the relationship in the data. Other analyses establish a maximum effect at tone precedence of 4 sec., with no effect for the backward conditions of −4 and −2 sec. precedence or for the long trace condition of 16 sec. precedence. Whether SP occurs at the backward condition of −1 sec. precedence and the trace condition of 8 sec. precedence is dubious. These two conditions represent the extremes between which fall the time relations during preconditioning that are effective in producing SP. A quadratic function with a maximum around 4 sec. precedence and zero at the backward condition of approximately −1-sec. precedence and the forward condition of approximately 8 sec. precedence may be reasonable, but is not clearly supported by the data of the experiment.

REFERENCES

BROGDEN, W. J. Sensory preconditioning. *J. exp. Psychol.*, 1939, 25, 323–332.
DUNCAN, R. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, 2, 171–189.
GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141–154.
HOFFELD, D. R., KENDALL, S. B., THOMPSON, R. F., & BROGDEN, W. J. Effect of amount of preconditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1960, 59, 198–204.
HOFFELD, D. R., THOMPSON, R. F., & BROGDEN, W. J. Effect of stimuli time relations during preconditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1958, 56, 437–442.
SILVER, C., & MEYER, D. R. Temporal factors in sensory preconditioning. *J. comp. physiol. Psychol.*, 1954, 47, 57–59.

(Received July 13, 1961)

---

# SUPPLEMENTARY REPORT: SEMANTIC GENERALIZATION IN PROBABILITY LEARNING

J. P. DAS

*Utkal University, Cuttack, India*

This experiment is planned to find out whether the introduction of semantic stimuli changes the course of generalization in probability learning. The present study is similar in design and procedure to one by Popper and Atkinson (1958). Both are based on predictions from a discrimination model and its application given in Estes and Burke (1955). Every trial begins with the presentation of a $T_1$ or a $T_2$ stimulus, the probability

of each being .50. Following a $T_1$ stimulus, the letter X occurs with a probability of .90 and Y with a probability of .10. But following a $T_2$ stimulus, X occurs with a probability of .70 for one group and .30 for another group; the corresponding probability of Y being .30 and .70, respectively. These probability schedules are identical with those of Groups II and IV of a related experiment by Atkinson, Bogartz, and Turner (1959).

TABLE 1

MEANS AND SDs OF X PREDICTIONS OVER 20 TRIALS FOLLOWING $T_1$ AND $T_2$

| Groups | | Trial Blocks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | | 4 | |
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| I | R90 | 13.53 | 2.53 | 17.30 | 1.80 | 18.40 | 2.18 | 18.73 | 1.61 |
| | D70 | 11.13 | 3.32 | 12.93 | 2.82 | 15.00 | 3.44 | 15.27 | 2.93 |
| II | R90 | 13.86 | 2.29 | 15.38 | 2.42 | 17.29 | 2.33 | 17.86 | 2.17 |
| | D30 | 8.62 | 1.89 | 7.29 | 3.04 | 4.62 | 2.34 | 3.33 | 2.71 |
| III | R90 | 14.54 | 4.52 | 17.38 | 2.27 | 17.85 | 1.70 | 17.92 | 2.23 |
| | N70 | 12.15 | 3.28 | 14.85 | 2.82 | 17.15 | 1.56 | 17.46 | 1.15 |
| IV | R90 | 13.38 | 2.27 | 14.38 | 3.20 | 15.85 | 3.30 | 16.46 | 3.22 |
| | N30 | 9.69 | 2.64 | 8.69 | 3.87 | 6.46 | 3.63 | 5.38 | 3.58 |

Note.—$t$ for R in I and III $= 1.04$, $df = 26$; for R in II and IV $= 1.33$, $df = 32$; for D and N in I and III $= 2.58$, $df = 26$, $P < .02$; for D and N in II and IV $= 1.71$; $df = 32$.

However, there is a major difference between the Popper and Atkinson experiment and the present one. Synonyms and antonyms have been used here in place of nonsense syllables as $T_1$ and $T_2$ stimuli.

*Method.*—The $T_1$ stimulus was an Indian word, RAJANI, meaning night. The two $T_2$ stimulus words, NIGHT and DAY, were synonymous and antonymous to RAJANI. When stimulus words were visually presented, they were written in the common script of $S$s, Oriya. There were four experimental groups: I, R90 D70 ($N=15$); II, R90-D30 ($N=21$); III, R90 N70 ($N=13$); and IV, R90-N30 ($N = 13$). R90-D70 stands for the group in which $T_1$ was RAJANI and probability of X after $T_1$ was .90, whereas $T_2$ was DAY and the probability of X after $T_2$ was .70. The other terms are similarly defined.

Each group of $S$s received 160 trials in all, divided into four 40-trial blocks. Within each block, the 20 $T_1$ and 20 $T_2$ trials were randomly distributed, and in the $T_1$ and $T_2$ blocks, the given proportions of X and Y were maintained. For example in Group I, out of the 20 $T_1$ cards, the combination RAJANI-X was found in 18 cards whereas the remaining 2 cards had RAJANI-Y. Similarly the 20 $T_2$ cards had 14 cards of DAY-X and 6 of DAY-Y.

The $S$s, college students, had four strips of paper for the four blocks of trials. On each strip they found two columns, marked with $T_1$ and $T_2$ stimuli at the top. Their task was to guess whether X or Y would come and record it under the appropriate columns immediately after $E$ read out the $T_1$ or $T_2$ stimulus. Then $E$ exposed the card, and at the same time read out whatever was written following the $T_1$ or $T_2$ stimulus. Each trial took about 10 sec.: 3 sec. to read out the stimulus word, 4 sec. allowed to $S$s for putting down their guesses, and 3 sec. to expose the card to $S$s and begin the next trial. The $S$s covered their previous choices by folding the recorded portion of the paper after each trial.

*Results.*—A summary of the findings is given in Table 1. The $S$s' predictions are expected to vary on two accounts—semantic relation between $T_1$ and $T_2$, and similarity between the probabilities of X after $T_1$ and $T_2$. Groups I and III, and Groups II and IV therefore, may be meaningfully compared. The final proportion of R choices in Group I should be higher than the same in Group III because there will be a greater generalization between the synonyms R and N than between the antonyms R and D. Similarly, the mean prediction of R in the final trial block for Group IV will be less than that for Group II, whereas the reverse would be true for N and D. The series of $t$'s computed shows that although the mean differences are in the expected direction, only the difference between N70 and D70 reached at least the .05 level of significance. This result may be interpreted as offering limited support to the hypothesis that semantic stimuli influence probability generalization.

REFERENCES

ATKINSON, R. C., BOGARTZ, H. W., & TURNER, R. N. Discrimination learning with probabilistic reinforcement schedules. *J. exp. Psychol.*, 1959, 57, 349–350.
ESTES, W. K., & BURKE, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, 50, 81–88.
PROPPER, J., & ATKINSON, R. C. Discrimination learning in a verbal conditioning situation. *J. exp. Psychol.*, 1958, 56, 21–25.

# THE EFFECTS OF DIFFERENTIAL VISUAL STIMULATION AFTER INDUCTION OF VISUAL AFTEREFFECTS [1]

HERBERT L. PICK, JR., MAVIS HETHERINGTON, AND ROLAND BELKNAPP

*University of Wisconsin*

The relationship between Gibson's negative aftereffect and the Köhler-Wallach figural aftereffect has been the subject of controversy.

Gibson (1933) found that a curved line, when perceived for a period of time, becomes phenomenally less curved than it was originally, and that after such an inspection period an objectively straight line appears curved in the opposite direction from the inspection curve. These negative aftereffects were attributed to a process similar to sensory adaptation in which curvature perceptions when long continued tend to approach the norm of a straight line. Gibson regards the straight line as a neutral from which other lines deviate. Thus a straight line will serve as an anchoring point in the perception of curved lines. A frequent condition of the environment tends to become a norm of the phenomenal world and new stimuli are perceived in relation to it.

Köhler and Wallach (1944) in their studies of figural aftereffects postulated an electrical field process in the visual cortex which satiates the cortex in the immediate area of the cortical representation of the inspection figure.

This satiation results in increased resistance to further stimulation in this area and to displacement of the cortical representation to neighboring regions upon subsequent stimulation. Köhler and Wallach proposed that Gibson's curved line effect could be adequately explained by their satiation theory, if the test line were considered to be displaced from the satiation area of the inspection curve.

Although Osgood and Heyer (1952) suggest an alternative mechanism to explain figural aftereffects, both their theory and the Köhler-Wallach theory are based upon the same sort of physiological satiation by prolonged inspection of contours.

Several recent attempts have been made to differentiate experimentally the Gibson negative aftereffect from the Köhler-Wallach figural aftereffect. Sagora and Oyama (1957) in their survey of studies of figural aftereffects in Japan cited considerable evidence to indicate that the curved line effect could not be explained by the Köhler-Wallach theory. Bergman and Gibson (1959), using a slanted textured surface as a stimulus, demonstrated that one type of negative aftereffect could not be explained on the basis of satiation and contour displacement. Gibson (1959a) has argued on logical grounds that the two kinds of aftereffect cannot possibly be the same, and has recently (Gibson, 1959b, pp. 489–491) restated the normalization hypothesis.

Such evidence as that reported by Sagora and Oyama (1957) and Berg-

man and Gibson (1959) suggest that different processes are involved in the two types of aftereffect. Further investigation seemed warranted to elucidate the nature of these differences and to clarify the theoretical bases of the phenomena.

The Köhler-Wallach theory suggests a gradual dissipation with time of the differential satiation in the visual cortex. If the dissipation is spontaneous, and if satiation explains both contour-displacement and curvature-straightening, it could be predicted that with homogeneous stimulation (e.g., a "ganzfeld") interpolated between the viewing of the inspection figure and of the test figure, both types of visual aftereffects should decrease. The Gibson theory, on the other hand, suggests that the normalization of a curved line would not dissipate during an afterperiod of homogeneous stimulation since the aftereffect represents simply a re-adaptation to the conditions of stimulation prevailing before the experiment. If the effects of homogeneous stimulation on the two types of visual aftereffects differ, the distinctiveness of the two phenomena will have been demonstrated and the adequacy of either theory to explain such differential results can then be evaluated.

## METHOD

In order to test the above predictions three different postinspection visual experiences were introduced following the induction of a Köhler-Wallach aftereffect and a Gibson curved line aftereffect, and their influence upon the aftereffects was then measured.

*Subjects.*—The Ss for the Köhler-Wallach figural aftereffect were 36 summer school students registered in introductory psychology courses; Ss for the Gibson negative aftereffect were 30 summer school students enrolled in introductory psychology or education courses.

*Apparatus.*—The apparatus for the Köhler-Wallach effect was a modified Dodge type

tachistoscope in which the inspection (I) figure was shown binocularly at a distance of 58 cm. from the eye. The test (T) figures were exposed at an equal distance. In both cases the visual field was 19.7 cm. square. The I figure consisted of a circle 5.5 cm. in diameter which was placed to the left of a fixation point in the center of the visual field. The width of the line defining the circumference of the circle was 2 mm. Each T figure was drawn on white cardboard and consisted of two circles placed on opposite sides and equally distant from a fixation point. The width of the line in the T circles was .75 mm. In half the test trials the circle on the left was 7.6 cm. in diameter and would tend to be expanded outward to produce an aftereffect of increased size. These circles were always concentric with the original I figure. When the left circle was 7.6 cm. in diameter, the circle on the right was 7.6, 7.8, 8.0, 8.2, or 8.4 cm. in diameter. In the other half of the test trials the circle on the left was 4.0 cm. in diameter and would tend to be compressed inward to produce an aftereffect of decreased size. When the circle on the left of the T figure was 4.0 cm. in diameter the circle on the right was 3.6, 3.7, 3.8, 3.9, or 4.0 cm. in diameter. In both cases the size of the right circle was varied randomly from trial to trial. The use of two different size circles on the left was necessitated by the possibility of response bias occurring if the figural aftereffects always occurred in the same direction. A measure of the magnitude of the figural aftereffect could be determined by noting size judgments of circles on the right in relation to circles on the left. The method of obtaining this measure is discussed in the section on scoring of performance.

The Gibson visual aftereffect was produced by having Ss fixate an I figure consisting of a black curved line 3 mm. in width and convex to the left drawn on white cardboard. In order to eliminate straight reference lines, the I line was presented through a circular aperture 30 cm. from the eye. The diameter of the aperture was 25.5 cm. The curved line extended the full length of the aperture and was bowed 2 cm. at its center. A series of five T lines was constructed and presented in the same manner. The T lines, 3 mm. in thickness, were also bowed to the left by 0.0, 0.1, 0.2, 0.3, and 0.5 cm. The T lines were presented in a random order. The magnitude of the aftereffect could be determined by noting which of these lines were perceived as straight.

The brightness of the larger surfaces (walls,

table-top) to which $S$ was exposed under conditions of normal stimulation varied from 4.0 to 4.4 ft-c as measured by a Macbeth illuminometer. The goggles for the ganzfeld condition passed approximately 25% of the illumination. The brightness of the I and T fields for the Köhler-Wallach figural aftereffects was 2.7 ft-c. The brightness of the I and T field for the Gibson aftereffects was 3.2 ft-c.

*Procedure.*—Hammer (1949), in the only published study of the dissipation of figural aftereffects, found that they decreased to zero in 150 sec. Pilot studies, using the conditions outlined above, indicated that with 5-min. I periods and subsequent normal visual conditions the Köhler-Wallach aftereffect diminished to zero in approximately 30 min.; the Gibson effect in 15 min. These time intervals were consequently used for the postinduction visual exposure periods in the respective experiments. Each experiment consisted of five steps in the following order: an initial control test with the T figures, a 5-min. fixation of the I figure, an immediate test of the magnitude of the aftereffect, a postinspection period in one of three visual conditions, and a final test of the magnitude of the aftereffect.

In the Gibson procedure the control test consisted in making judgments of the direction of curvature of the five curved test figures convex to the left plus two curved figures convex to the right. The latter curved lines were included to accustom $S$ to perceive lines curved in both directions. These lines were presented in random order. This procedure required approximately 20 sec. depending on the speed of judgment of $S$. During the 5-min. I period which followed the control test, $S$ was instructed to run his eyes slowly up and down the middle portion of the I line. The immediate test of the magnitude of the aftereffect was the same as the control test except for the omission of the curves convex to the right. One-third of the $S$s then received each of the following postinspection conditions for 15 min.: Cond. I, normal stimulation, $S$ looked around the room; Cond. II, homogeneous light stimulation, $S$'s vision was limited to a homogeneous ganzfeld, produced by goggles which covered each eye with a concave section of a translucent ping pong ball; Cond. III, homogeneous lack of stimulation, $S$ had a black blindfold over open eyes. A final test of the magnitude of the aftereffect was administered as before.

In the Köhler-Wallach procedure the initial control test consisted of viewing the 10 test cards with two black circles on them. Cards were presented in random order at 4-sec. intervals. The $S$ was instructed to fixate on the cross between the circles and report which of the circles was larger or if they were equal in size. During the 5-min. I period $S$ was directed to fixate on the cross on the I figure. The subsequent test utilized the same procedure and stimuli as the control test. One-third of the $S$s were then exposed for 30 min. to each of the three postinspection conditions described above and finally the third test of the magnitude of the figural aftereffect was administered in the same manner as the previous ones.

*Scoring of performance.*—The scoring procedure for the Köhler-Wallach aftereffect was as follows: An arbitrary scoring system assigned numerical values to the size of the figural aftereffect. If the 7.6-cm. circle on the right was judged to be equal to the 7.6-cm. circle on the left (standard), a score of zero was assigned; if it was judged smaller (this being evidence of a figural aftereffect), a score of +1 was assigned. If the 7.8-cm. circle on the right was judged smaller than the standard, a score of +2 was assigned. Thus the weights were increased by integral units as the magnitude of figural aftereffect increased. An identical procedure was used with the 4.0-cm. standard circle, but since the effect of the inspection was to make the 4.0-cm. standard appear smaller, evidence of the visual aftereffect was a tendency of the circle on the right to appear larger.

A similar procedure was used to score the magnitude of the aftereffect of the Gibson type.

## RESULTS AND DISCUSSION

With the scoring procedures a value was computed for each $S$ for the control test, the initial aftereffect test, and the final aftereffect test. These were computed by totaling the number of points obtained by $S$ on all the test stimuli. The average values obtained and their $SD$s are shown in Table 1. In each experiment the values obtained for each of the three groups could be compared. The differences between the magnitude of the aftereffect (immediately after inspection) and the final magnitude (after the interpolated visual experience) were analyzed with each $S$ serving as

## TABLE 1

### AVERAGE MAGNITUDES AND *SD*s OF AFTEREFFECT SCORES

| Test | Köhler-Wallach | | | | | | Gibson | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | Normal | | Blindfold | | Ganzfeld | | Normal | | Blindfold | | Ganzfeld | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Control | 2.8 | 3.5 | 2.6 | 3.8 | 2.0 | 2.1 | 0.1 | 0.3 | 0.1 | 0.7 | 0.2 | 0.4 |
| Initial aftereffect | 6.6 | 4.2 | 7.2 | 5.4 | 5.7 | 3.3 | 3.0 | 2.0 | 2.8 | 1.9 | 4.4 | 2.2 |
| Final aftereffect | 1.8 | 4.1 | 5.6 | 4.2 | 8.1 | 3.0 | 0.9 | 1.0 | 0.6 | 0.5 | 0.6 | 0.5 |

his own baseline. Negative differences occurred where the final aftereffect test resulted in a lower score than the immediate postinspection test, i.e., a decrease of the aftereffect over time. Conversely, positive differences indicated an increase of aftereffect over time.

The results shown in Table 1 suggest that with the Köhler-Wallach experiment the ganzfeld condition serves to enhance the aftereffect, and although the blindfold condition does not enhance the aftereffect it does reduce the amount of decay. The greatest decrease in aftereffect occurred with the normal vision group. With the Gibson aftereffect the greatest decay occurred under the ganzfeld condition. These findings are verified when the data are subjected to analysis of variance.

With the Köhler-Wallach effect the three groups are significantly different from each other ($P < .01$) and the ganzfeld group shows a significant enhancement effect ($P < .05$), i.e., an increase in magnitude of aftereffect over the 30-min. period. A significant difference ($P < .025$) between groups found with the Gibson aftereffect is apparently due to the fact that the ganzfeld group had somewhat inflated values in the initial aftereffect test. In any case no significant differences in final level of aftereffect occurred with the Gibson aftereffect, i.e., the aftereffect had vanished completely under all three visual conditions.

Since conditions of homogeneous stimulation effect these two visual aftereffects in radically different ways, it is strongly implied that the two phenomena are basically different.

The results are, in general, opposite to what might be expected if each theory predicted the outcome for its own type of aftereffect. The inference from the Köhler-Wallach theory was that a diminution of the aftereffect would occur following conditions of homogeneous stimulation. However, the Köhler-Wallach aftereffect was enhanced by the ganzfeld situation. Gibson's normalization theory would predict a maintenance of the curved line aftereffect under homogeneous stimulation but instead this effect had essentially vanished.

The nature of the perceptual system which might account for the paradoxical enhancement or the maintenance of the Köhler-Wallach effect warrants further investigation. It may be related to the higher susceptibility to figural aftereffects found after a period of sensory deprivation (Doane, Manatoo, Heron, & Scott, 1959). In the present study, if the ganzfeld and blindfold conditions were considered conditions of sensory deprivation it would suggest that stimulation just prior to such deprivation has a particularly strong effect. If the homogeneous conditions are related to sensory deprivation the differences obtained under the two conditions require further investigation.

## SUMMARY

The effects of three conditions of post-stimulation on a Köhler-Wallach figural after-effect and a Gibson negative aftereffect were investigated. Condition I was normal stimulation, obtained by looking around the room; Cond. II was homogeneous lack of stimulation, obtained by wearing a black blindfold; Cond. III was homogeneous light stimulation obtained by exposure to a "ganzfeld." The Gibson aftereffect decreased normally under all three conditions. The ganzfeld enhanced the Köhler-Wallach aftereffect, the blindfold retarded the decrease in the aftereffect, and looking around the room permitted the normal disappearance of the figural aftereffect.

## REFERENCES

BERGMAN, R., & GIBSON, J. J. The negative after-effect of the perception of surface slanted in the third dimension. *Amer. J. Psychol.*, 1959, **72**, 364–374.

DOANE, B. K., MANATOO, W., HERON, W., & SCOTT, T. H. Changes in perceptual function after isolation. *Canad. J. Psychol.*, 1959, **13**, 210–219.

GIBSON, J. J. Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.

GIBSON, J. J. After-effects: Figural and negative. *Contemp. Psychol.*, 1959, **4**, 294–295. (a)

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. I. New York: McGraw-Hill, 1959. Pp. 456–501. (b)

HAMMER, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62**, 337–354.

KÖHLER, W., & WALLACH, H. Figural after-effects. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

SAGORA, M., & OYAMA, T. Experimental studies on figural after-effects in Japan. *Psychol. Bull.*, 1957, **54**, 327–338.

# EFFECTS OF PROBABLE OUTCOME INFORMATION ON TWO-CHOICE LEARNING

## RICHARD C. NIES [1]

*University of California, Los Angeles*

The impact of statistical learning theory (Bush & Mosteller, 1955; Estes, 1950) has focused attention upon probability learning in the two-choice situation (Goodnow, 1958; Humphreys, 1939). Although success is maximized by consistent choice of the more likely alternative, *S*s typically approximate the actual proportions of the reinforcement schedule (Hake, 1955; Jenkins & Stanley, 1951). However, when the task is presented in a gambling context (i.e., total correct choices are maximized), *S*s tend to predict the more frequent event at a significantly higher proportion of the trials (Goodnow, 1955; Siegel & Goldstein, 1959). The present study was designed to further explore the conditions which maximize success by manipulating information about the probability of the events. Although responses in two-choice situations appear to be relatively independent of experimental instructions (Anderson & Grant, 1957; Neimark & Shuford, 1959) and performance information (Das, 1961), Koehler (1961) was able to predict mean terminal response rates by varying instructions which dealt with how *S*s should consider the nonreinforced trials in a two-choice contingent partial reinforcement situation. In view of this finding, the stability of behavior to the manipulative effects of differential information in such situations is to be questioned. It was

therefore hypothesized that specific information about the probabilities of two alternative outcomes would produce a shift toward consistent selection of the more likely event.

## METHOD

*Apparatus.*—A box containing 100 marbles, 70 of one color and 30 of another color, was mounted so that the marbles could be mixed by turning a crank protruding from one end. A trough at the other end received 1 marble whenever the box was tilted forward, the marble rolling back into the box when the apparatus was returned to its normal position.

*Subjects.*—Eight experimental and four control sessions were held, with groups of 16 *S*s drawn for each session from the course in introductory psychology at the University of California, Los Angeles.

*Procedure.*—The following instructions were read aloud to all *S*s in the experimental groups:

This experiment is designed to study human guessing habits. In this box are a number of blue and red marbles (shown briefly). For each trial, I will shake the box thoroughly and allow one marble to roll into the trough (demonstrated). While I am shaking the box I will count to three (demonstrated), and by the time I say "three," you write down on the response sheet, which will be given you shortly, whether you think a blue marble or a red marble will roll out. After you have written your choice, I will allow a marble to roll out and you will then be told its color. Your task is to get as many correct predictions as you can.

To assure the 70-30 ratio of blue to red marbles, a secret sheet was prepared which maintained this schedule randomly in blocks of 50 (35 blue marbles to 15 red marbles) over 250 trials. The random sequences were chosen with no constraints from a table of random numbers according to their probabilities. Although an assistant read the color of the marble for each trial from this prepared

schedule, Ss were led to believe that he was reporting the color of the marble that actually rolled out. This procedure insured that the different groups were exposed to identical sequences.

The Ss in each experimental session were randomly divided into four equal subgroups, differentiated with respect to the additional information printed at the top of their response sheets. The No-Information group had no additional information. The Pattern group had the following sentence added: "It has been found in experiments of this kind that the marbles roll out in definite patterns." The Ratio group had the statement: "There are 100 marbles in the box: 70 are blue and 30 are red." And the Ratio-Explanation group was informed that: "There are 100 marbles in the box: 70 are blue and 30 are red. This means that the chances of a blue marble rolling out are 7 out of 10, and the chances of a red marble rolling out are 3 out of 10. Since each marble that rolls out is returned to the box, these odds will be the same for each trial. Furthermore, since the box is shaken thoroughly each time, there can be no fixed pattern in which the marbles roll out."

For all conditions, S was to indicate his prediction by recording either an "R" or a "B" on each trial.

To control for color preferences and sequence peculiarities, half the experimental Ss were run with colors reversed (i.e., 70 reds to 30 blues), and half were run using a different 70-30 sequence. To control for unique group differences, each of these four experimental combinations was replicated with a second group. The result was a 2 × 2 × 4 factorial design (Color × Sequence × Information) replicated with a second set of four experimental groups.

Four control groups (two groups for each reinforcement sequences under the .70-red condition only) were also run to provide a baseline against which the effects of the marble box could be assessed. They were read the following instructions:

This experiment is designed to study human guessing habits. For each trial, I will call out one of two colors—blue or red. Before I announce the color, I will count to three (demonstrated). While I am counting, you write down on the response sheet, which will be given you shortly, whether you think I will call out blue or red. After you have written your choice, I will then announce the color. Your task is to get as many correct predictions as you can.
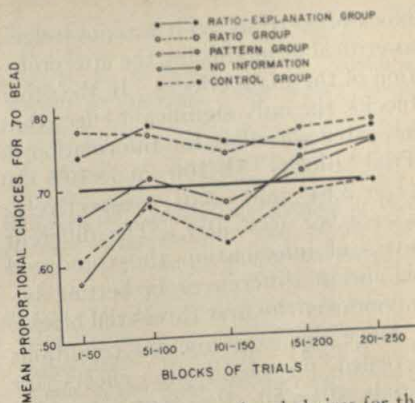


FIG. 1. Mean proportional choices for the .70 marble by trial blocks under various levels of information.

## RESULTS AND DISCUSSION

Figure 1 shows the mean proportion of choices for the more likely alternative over each successive block of 50 trials. Separate analyses of variance were performed for each block of 50 trials, using the mean proportions as raw scores. Table 1 summarizes the analysis for the first 50 trials, showing the significant effect of Information and also of the interaction between Information and Sequence. This interaction was found only in the first

TABLE 1

ANALYSIS OF VARIANCE FOR TRIAL BLOCKS 1–50 FOR EXPERIMENTAL GROUPS

| Source | df | MS | F |
|---|---|---|---|
| Instructions (I) | 3 | 639.80 | 14.55** |
| Color (C) | 1 | 54.82 | 1.44 |
| Sequence (S) | 1 | 2.32 | .06 |
| I × C | 3 | 5.55 | .12 |
| I × S | 3 | 156.18 | 3.53* |
| C × S | 1 | 19.33 | .51 |
| I × C × S | 3 | 40.88 | .92 |
| Groups within treatments[a] | 4 | 38.10 | .86 |
| Pooled Groups × Instructions | 12 | 51.91 | 1.17 |
| Ss within groups[b] | 96 | 44.28 | |

[a] Error term for C, S, and C × S.
[b] Error terms for I, I × C, I × S, I × C × S, Groups within treatments, and Pooled Groups × I.
 * Significant at .05 level.
** Significant at .001 level.

block of 50 trials, so it was not judged as critical in modifying the interpretation of the main effects. In the other blocks, the only significant effect that appeared was that for Information in Trial Blocks 51–100 and 101–150 ($F = 3.64$ and $3.61$, respectively; $df = 3/96$; $P < .05$). The different levels of information thus produced significant differences in betting behavior over the first three-trial blocks.

Since the experimental conditions revealed no significant effects for Trials 201–250, the scores for this block were pooled (using the .70-red condition only) and tested against the control scores for the same block. The difference was highly significant ($F = 16.33$; $df = 1/100$; $P < .001$).

These analyses support the following conclusions: (a) The information about probable event outcomes affect $Ss$' responses during the early blocks of trials (up to 150 trials), but shows no significant effect in later blocks. These early differences appear to be relatively independent of color preferences, sequence peculiarities, and group differences. (b) While the control groups (no marbles) just reach the probability matching level on the last block of trials, the experimental groups (marbles) exceed this level for all the trial blocks. Relevant to this finding is the proposal by Flood (1954) and Rubinstein (1959) that the awareness of randomness in a two-choice probabilistic outcome situation tends to elevate $S$'s response predictions, the impossibility of a complete solution increasing $S$'s caution against betting on the less likely alternative. Randomness was made explicit in the present study by the use of a box in which marbles were thoroughly mixed in $Ss$' presence before each prediction was made. In the control groups with no marbles, $Ss$ displayed typical probability matching behavior.

While consistent with the interpretation made by Flood and Rubinstein, the present experiment advances their argument by demonstrating an early facilitation of optimal betting through the introduction of specific information relevant to the probable outcome of the events. While the presence of the marble box can account for the higher response level of the experimental groups, presumably through its contribution to the random appearance of the events, the perception of randomness was hastened by the critical information supplied to Ratio and Ratio-Explanation groups. When information is effective in determining behavior in a two-choice, uncertain outcome situation, its effectiveness may be based on its contribution to the perception of randomness. The fact that the instructions used by Anderson and Grant (1957) and Neimark and Shuford (1959) did not elevate $Ss$' response levels may result from a failure to make explicit the randomness and impossibility of a complete solution.

Although the experimental conditions elevated the proportion of correct anticipations, only 4 $Ss$ (all from the Ratio-Explanation group) learned to bet 100% on the more likely alternative. Two suggestions for this relative lack of optimal betting find support in the present research: (a) The $Ss$ seem surprisingly expectant of patterns and systems when there is any challenge of a problem to be solved (Goodnow, 1958). This is consistent with the responses to a questionnaire given at the conclusion of the present experiment where 75 of the 128 experimental $Ss$ reported looking for a pattern whereas only 32 had been told (falsely) that there was one. Also, as shown in Fig. 1, the betting curve for the Pattern group is fairly close to that of the No-Information group. Moreover, while patterns may be expected from the problem solving nature of the task, certain characteristics of the data support it. In contrast to the method of constrained randomization used by Edwards (1961; Lindman & Edwards, 1961) and Nicks (1959), the method used in this research involved an unrestrained randomization, resulting in a distribution of lengths of homogeneous outcome runs which included too many short runs and not enough long runs. Thus there were

patterns in the outcome sequence which allowed for the gambler's fallacy. The point to be made here is that this reduction from optimal betting is not "irrational" in terms of the Ss' set to expect patterns. (b) Furthermore, the Ss prefer the occasional success of guessing the unlikely alternative to the monotony of repetitiously predicting the more certain event where no tangible inducements are offered to maximize their correct predictions (Goodnow, 1955; Siegel & Goldstein, 1959). Thus, there is a second source of reinforcement at work, the utility of correctly predicting the occurrence of the less frequent event, which subtracts from maximum gain responding (Brackbill, Kappy, & Starr, 1962). This also is consistent with the questionnaire responses; 77 of the experimental Ss reported that they knew the wisest procedure would be to bet consistently on the more likely alternative.

The data, thus, suggest a compromise between the expedience of betting 100% for the more likely alternative (because of the "chance" or "random" structure made patent by the experimental apparatus) and the challenge of betting 70% for the more likely alternative (i.e., trying out different patterns in an effort to "beat" the game).

## Summary

The experiment investigated the effect of different levels of probability information on response frequencies in a random, two-choice situation with unequal event probabilities. This information was demonstrated to have differential effects during the early trials. In addition, it was found that Ss reached a significantly higher response level when the outcome of a trial appeared to depend upon the chance drawing of a marble from a box than when the marble box was absent. These findings were interpreted in terms of the perception of "randomness."

## References

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. J. exp. Psychol., 1957, 54, 305–317.

BRACKBILL, Y., KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability

learning. J. exp. Psychol., 1962, 63, 32–35.

BUSH, R. R., & MOSTELLER, F. Stochastic models for learning. New York: Wiley, 1955.

DAS, J. P. Mathematical solution in the acquisition of a verbal CR. J. exp. Psychol., 1961, 61, 376–378.

EDWARDS, W. Probability learning in 1000 trials. J. exp. Psychol., 1961, 62, 385–394.

ESTES, W. K. Toward a statistical theory of learning. Psychol. Rev., 1950, 57, 94–107.

FLOOD, M. M. Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), Decision processes. New York: Wiley, 1954. Pp. 287–299.

GOODNOW, J. J. Determinants of choice distributions in two-choice probability situations. Amer. J. Psychol., 1955, 68, 106–116.

GOODNOW, J. J. A review of studies on probable events. Aust. J. Psychol., 1958, 10, 111–125.

HAKE, H. W. The perception of frequency of occurrence and the development of "expectancy" in human experimental subjects. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Ill.: Free Press, 1955. Pp. 257–274.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. J. exp. Psychol., 1939, 25, 294–301.

JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: A review and a critique. Psychol. Bull., 1951, 41, 291–297.

KOEHLER, J., JR. Role of instructions in two-choice verbal conditioning with contingent partial reinforcement. J. exp. Psychol., 1961, 62, 122–125.

LINDMAN, H., & EDWARDS, W. Supplementary report: Unlearning the gambler's fallacy. J. exp. Psychol., 1961, 62, 630.

NEIMARK, E. D., & SHUFORD, E. H. Comparison of predictions and estimates in a probability learning situation. J. exp. Psychol., 1959, 57, 294–298.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. J. exp. Psychol., 1959, 57, 105–114.

RUBINSTEIN, I. Some factors in probability matching. J. exp. Psychol., 1959, 57, 413–416.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. J. exp. Psychol., 1959, 57, 37–42.

# COLOR CODING AND VISUAL SEARCH [1]

SIDNEY L. SMITH

*The MITRE Corporation, Bedford, Massachusetts*

There is relatively little available data on the effectiveness of color as a coding dimension for information transmission in visual displays. Some studies have been published exploring the limitations of our ability to identify different hues on the basis of absolute discrimination. They suggest that under good viewing conditions 5 to 8 colors can be distinguished reliably (Conover, 1959; Conover & Kraft, 1958; Eriksen & Hake, 1955), or perhaps even as many as 9 to 12 under optimal conditions (Chapanis & Halsey, 1956; Halsey & Chapanis, 1951).

Whichever number we accept, this means that color as a coding dimension can be used to distinguish only a relatively small number of categories, as compared with shape coding (symbology) which carries the bulk of information in visual displays. However, within this limitation, there is evidence that color is better than shape in tasks which involve locating displayed data (Christner & Ray, 1961; Hitt, 1961).

The effectiveness of color coding for locating particular displayed items is related, by extension, to its use for providing visual separability *among* data classes. Results on this question have been reported by Green and Anderson (1956). They examined the degree to which color coding permitted visual separability of displayed two-

digit numbers, as measured by decreases in average visual search time when the color of a "target" number was known beforehand (relevant) as compared with other occasions when it was not known (nonrelevant). They (Green & Anderson, 1956) summarized their conclusions as follows:

> When *O*s know the color of the target, the search time is approximately proportional to the number of symbols of the target's color. There is also a slight increment in search time due to the presence of the wrong-colored targets. When *O*s do not know the target's color, search time depends primarily on the total number of symbols on the display. However, search times are slightly longer for multicolored displays than for comparable single-colored displays (p. 24).

In an attempt to confirm these results, this present experimental study was conducted. This study accepts the basic premise of Green and Anderson that visual search time is a fundamental measure of the potential value of display color coding. It simply expanded their model to include a greater range of displayed densities, more displayed colors, both light and dark display backgrounds, and certain modified techniques of display presentation.

## PROCEDURE

Twelve *S*s participated in this study, 11 men and 1 woman. Preliminary testing using the American Optical Company H-R-R pseudoisochromatic plates confirmed that all *S*s had normal color vision. In the course of the study, each *S* made a total of 300 visual searches, using a variety of different displays. This required several experimental sessions for each *S*. Individual *S*s worked for no longer than 1 hr. at a time, and for no more than 2 hr. per day.

The displays consisted of varying arrays of three-digit numbers. These numbers were

randomly placed in a square field, which can be imagined as comprising 13 columns and 27 rows, for a total of 351 possible positions. The numbers themselves were chosen randomly from the 1000 possibilities (000 through 999) with certain restrictions: the numbers on each particular display were all different; they were unique in terms of their first two digits; their third digits represented an equal sample of each of the 10 possibilities (0 through 9). In the case of multicolored displays, the particular color of each three-digit number was also chosen at random, with the restriction that all colors were equally represented.

The displays were made up as 2 × 2 in. color slides, and presented to Ss by rear projection on the screen of an experimental console. The dimensions of the digits as projected on this screen were $\frac{1}{3}$ in. high × $\frac{1}{8}$ in. wide. Viewing distance was about 18 in. The overall display field as projected was 12 in. square.

The colors used on the various displays were red, green, blue, orange, and either white (on slides with a black background) or black (on a white background). Visual matches of the colors as projected with standard colors resulted in agreement by 2 Os on the descriptive specifications shown in Table 1.

Because rear projection was used, a moderately high ambient illumination was maintained, by diffuse overhead lighting from dimmed fluorescents: over .5 ft-c as measured at the experimental console.

The experimental routine for S began when E signaled to him the first two digits of the target number, the number he was to find on the display. These two digits were displayed on an auxiliary panel. The S also saw on this panel either a colored indicator showing what color the target number would be, or else a statement "color unknown." The E then exposed the slide on the screen in front of S and started a clock. The S searched for the target number, indicated when he had found

TABLE 1

DISPLAY COLORS USED (MUNSELL NOTATION)

| Display Color | White Background | Black Background |
|---|---|---|
| Red | 2.5 R 5/10 | 5 R 5/12 |
| Green | 7.5 GY 8/8 | 5 GY 7/8 |
| Blue | 2.5 PB 6/8 | 2.5 PB 6/8 |
| Orange | 2.5 YR 7/10 | 2.5 YR 7/10 |
| Black/White | 7.5 P 3/4 | N 9/0 |

TABLE 2

DISPLAYS USED OF EACH TYPE

| Number of Colors Displayed | Number of Displayed Items | | | | |
|---|---|---|---|---|---|
| | 20 | 40 | 60 | 80 | 100 |
| 1 | 5 | 5 | 5 | 5 | 5 |
| 2 | | 10 | | | |
| 3 | | | 10 | | |
| 4 | | | | 5 | |
| 5 | 5 | 5 | 5 | 5 | 5 |

it by pushing one of 10 numbered buttons, corresponding to the third digit of the target number. If this response was correct, the clock stopped, a chime sounded, and the screen went blank. The E recorded the time, and then set up the auxiliary display panel to begin the next trial. In the very rare case when S pushed the wrong response button, a loud buzzer sounded, and the trial was run again later in the experimental series.

The displays were presented in a series of 150 slides, half with white backgrounds and an identical group with black backgrounds. Each group of 75 slides contained both single- and multicolored displays with display densities varying from 20 to 100 three-digit numbers (which will be called "items" in the next few paragraphs to avoid confusion). The types of displays used are summarized in Table 2. The entries in the matrix represent the number of different displays used.

There are really three sets of displays represented in the matrix. The upper row consists of single-colored displays of increasing display density, with 5 displays (one of each color used) at each density level. The lower row consists of a comparison set of displays, each with items of all five colors on it. It was arbitrarily decided to use 5 displays at each density level, to match the first set. The diagonal set of the matrix represents those displays where an increasing number of displayed colors is associated with increasing display density, with the advantage that a constant number (20) of items of any particular color are displayed amid varying numbers of other-colored items: 10 two-colored displays were used, representing all the possible pairs of the five colors available; 10 three-colored displays represented all the possible triple combinations; and 5 four-colored displays represented the cases where each of the five available colors had been omitted.

The series of 150 displays described above

were presented in a random order, which in turn was randomized (in five blocks) differently for different $Ss$. Each $S$ went through this series twice in the course of the experiment, for a total of 300 searches. The first time through, for half of the displays (randomly chosen) he was told in advance the color of the target number, the number-to-be-searched-for. For the other half of the displays, the color of the target number was unknown to him. During his second run through the series, these conditions were reversed for each particular display—where the color of the target had been known before, it was now unknown, and vice versa.

In summary, visual search time data were obtained from 12 $Ss$, each viewing a series of 300 displays, which varied in display density, in number of colors used, in the particular color of the target, with either a white or black background, under conditions where $S$ either knew the color of the target number in advance or did'not.

## RESULTS AND DISCUSSION

Because of the relative complexity of the experimental design, no single statistical treatment would suffice to answer all the questions we might wish to raise. In fact, five separate analyses of variance were made, each using some portion of the data, in order to examine the effect of different combinations of the experimental variables.

The most extensive analysis of variance, and one basic to the subsequent data analysis, was carried out using individual search times obtained under all conditions involving either single-colored or five-colored displays. These data represent a factorial design of 12 $Ss$ by 2 conditions of prior knowledge (target color either known or unknown) by 2 types of display background by 5 degrees of display density by 2 types of display (either single-colored or five-colored) by 5 possible colors used for the target number. Taken together they comprise 2400 measures.

The data initially available for this

analysis were search times expressed in .01 sec. Because of the inherently high correlation between mean and variance in search time data, a log transform was used prior to the variance analysis computations. The computations themselves were carried out by an electronic computer, and sums of squares were obtained for the six experimental variables and their 57 various interaction terms, including the six-way residual. Following a procedure suggested by Edwards (1950) for the treatment of repeated measures obtained from the same $Ss$, all sums of squares representing interactions of $Ss$ with other experimental variables were pooled to form one residual representing the $Ss$ $\times$ Conditions interaction. The mean square of this interaction was used as the error term to test the significance of all others. (As it happens, if $Ss$ *had* been treated as a legitimate variable in the analysis, none of the interactions involving $Ss$ would have proved statistically significant.) Because of the large number of $F$ ratio comparisons made in this analysis, only significance levels of at least .001 were accepted as persuasive evidence of statistical reliability. The summarized results of this analysis are presented in Table 3.

We may note, first, that neither the particular target color used, nor the display background, nor any interaction term including these variables, had any statistically significant effect. We might have expected, for example, that some colors on a light background would be less legible than on a dark background, and hence more difficult to scan quickly, since it is clear that visual contrast is an important variable in legibility. However, the present data suggest that we need not expect any measurable effect related to different contrast ratios

TABLE 3

Analysis of Variance of Transformed Search Times for All Experimental Conditions Involving One- and Five-Colored Displays

| Source of Variance | df | MS | F |
|---|---|---|---|
| Ss | 11 | 0.708 | — |
| Conditions | 199 | — | — |
| Knowledge of target color (K) | 1 | 21.787 | 229.34* |
| Display background (B) | 1 | 0.301 | 3.17 |
| Display density (D) | 4 | 16.791 | 176.75* |
| Number of colors (N) | 1 | 14.956 | 157.43* |
| Color of target (C) | 4 | 0.063 | 0.66 |
| K × B | 1 | 0.015 | 0.16 |
| K × D | 4 | 0.619 | 6.52* |
| K × N | 1 | 15.549 | 163.67* |
| K × C | 4 | 0.106 | 1.12 |
| B × D | 4 | 0.071 | 0.75 |
| B × N | 1 | 0.000 | 0.00 |
| B × C | 4 | 0.095 | 0.98 |
| D × N | 4 | 0.129 | 1.36 |
| D × C | 16 | 0.129 | 1.36 |
| N × C | 4 | 0.292 | 3.06 |
| Higher-order interactions among conditions | 145 | 0.096 | —ᵃ |
| Ss × Conditions | 2189 | 0.095 | — |

ᵃ None was significant when tested individually.
* $P < .001$.

that are all well above some sort of legibility threshold.

Next, consider the marked effect on search time of increasing display density, confirmed by the statistical analysis and illustrated in varying degrees by Fig. 1–3. It is the nature of these data that there is greater inherent variability of the search time measure in those situations where the average search takes longer. Taking this into account, it seems that the average search time data closely approximate a direct linear relation with display density. This was also the case in the Green and Anderson study and has been noted before by other investigators (e.g., Green, McGill, & Jenkins, 1953).

If the "curves" in Fig. 1 are extrapolated backward to a display density of zero, they would intersect the ordinate axis at a value of about 1 sec., which presumably represents the simple button-pushing reaction time. This is confirmed by the observation that only one or two of the individual search time measures obtained in this

study were less than 1 sec., and those were smaller by only a slight margin. The same observation was made by Green and Anderson (1956) under similar circumstances. It should be pointed out that their reported search times were in general somewhat shorter than those in the present study. In part, this is because they reported geometric rather than arithmetic means. And, in part, this difference is probably attributable to the fact that their display field (12 × 16¾ in., viewed from a distance of 10 ft.) subtended a smaller visual angle than was the case in this present study. Hence, fewer eye movements were required to scan their displays.

The significant interaction between knowledge of target color, and number of colors displayed, indicates the potential value of color coding when the color is relevant to the search task. This is illustrated by the sizable difference between the curves in Fig. 1. If the color coding permitted absolute visual separability between classes of displayed items, then the average search times for these five-



FIVE - COLOR DISPLAYS
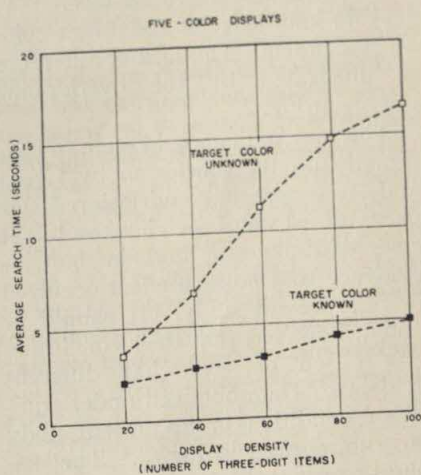
Fig. 1. Search time as a function of display density with knowledge of target color as a parameter.

color displays would be only one-fifth as great when the target color was known as when it was unknown. In actuality, the difference was not quite so great as this, but considerable nonetheless.

Comparable curves for single-colored displays showed no such difference, i.e., knowing the target color beforehand did not speed the subsequent visual search process. Indeed, there is no logical reason to expect that it would. However, to check this, a separate variance analysis was conducted, similar to that already described, but using only the search times for single-colored displays. The only significant variable in this treatment turned out to be display density. Prior knowledge of the target color made no difference.

A comparable variance analysis was run comparing one- and five-colored displays, but using only the search time data obtained when Ss had no prior knowledge of target color. Again, the only significant variable proved to be display density. In particular, number of colors displayed made no difference to Ss when they had no knowledge of the target color to guide them. This lack of difference is illustrated by the overlapping curves in Fig. 2.

This last finding would seem to bear directly on the conclusion by Green and Anderson that multicolored displays retard visual search when target color is unknown, and in fact, is evidence that no such effect occurred in this present study. It should be noted, however, that their conclusion was based on somewhat different evidence. Data obtained under equivalent conditions in this present study are summarized in Fig. 3. The displays considered here are those which always contained 20 items of the same color as the target number. For the
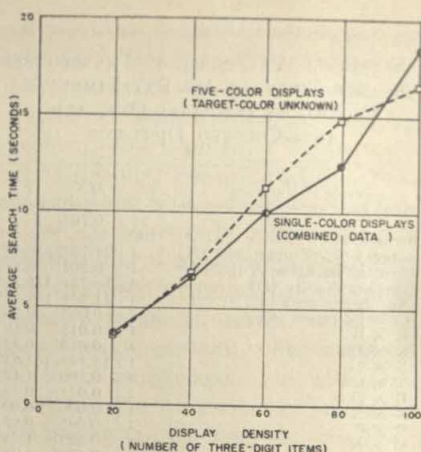


FIG. 2. Search time as a function of display density under conditions of nonrelevant color coding and of no color coding (single-colored displays).

displays with just one color, this is all they contained. For the two-colored displays, there were also 20 items of the second color present. For the three-colored displays, there were 20 items displayed for *each* of the two other (nontarget) colors—and so on, for the four- and five-colored displays.
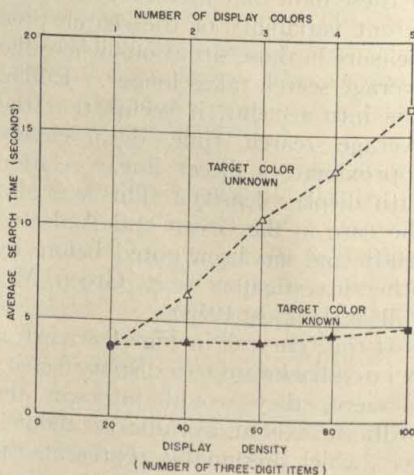


FIG. 3. Search time as a function of display density when target color is known and unknown and when there are 20 items of target color in all cases.

This comparison in Fig. 3, showing the considerably reduced search times when the target color was known to $Ss$, is essentially the comparison made by Green and Anderson in their report. It is true that the number of displayed colors is here confounded with differences in display density. But in the sense that each display contains just 20 target numbers amid varying amounts of "clutter," the comparison it permits is interesting.

To permit a direct check of the Green and Anderson (1956) results, a further analysis of variance was run, comparing search times for single-colored displays, of 40, 60, 80, and 100 items, with those for two-, three-, four-, and five-colored displays, of corresponding respective densities, under conditions of no prior knowledge of target color. Since neither display background nor target color had proved to be significant variables in preceding analyses, these were eliminated in this analysis by summing individual search times across these variables. Thus the data used consisted of the sum of 10 search times (5 target colors by 2 display backgrounds) under each combination of display density and number of colors. Actually, this is not entirely true, for in the case of the two-colored, 40-item displays, and the three-colored, 60-item displays, there were 20 search times available (10 target color/other color combinations by 2 display backgrounds). In these cases, half the available data were chosen randomly to use in the variance analysis. This total data subset, then, represents a factorial design of 12 $Ss$ by 4 degrees of display density by 2 types of displays—single-colored versus multicolored—for a total of 96 measures.

The results of this analysis are summarized in Table 4. The only

TABLE 4

ANALYSIS OF VARIANCE OF SUMMED SEARCH TIMES FOR CONDITIONS INVOLVING SINGLE- AND MULTICOLORED DISPLAYS WITH NO PRIOR KNOWLEDGE OF TARGET COLOR

| Source of Variance | $df$ | $MS$ | $F$ |
|---|---|---|---|
| $Ss$ | 11 | 9,351 | — |
| Conditions | 7 | — | — |
| Number of colors (N) | 1 | 3,449 | 0.71 |
| Display density (D) | 3 | 68,973 | 14.11* |
| N × D | 3 | 4,397 | 0.90 |
| $Ss$ × Conditions | 77 | 4,888 | — |

* $P < .001$.

variable with a statistically significant effect proves again to be display density. There is no reliable difference that can be attributed to number of displayed colors. If such a difference had been confirmed, it would have been in favor of the multicolored displays, which were searched in an average time of 5.9 sec., as compared with 6.5 sec. for the single-colored displays. Certainly we cannot conclude that multicolored displays were distracting in this present study, even though the target color was unknown.

Why did Green and Anderson obtain a different result? The most probable explanation lies in their choice of projection technique: luminous colored numbers displayed in a dark environment would provide a stimulus situation very conducive to perception of a depth illusion based on differences in either brightness or wave length. If their displayed numbers of different colors appeared to be in slightly different frontal planes, this might have led their $Ss$ to scan a multicolored display several times, a process that on the average would be somewhat slower than one systematic search of a single-colored display. In this present study, with rear projection and higher ambient illumination, the displayed symbols clearly lay on a single surface, and no depth effects were apparent.

The slight but regular increase in search time, when the target color was known, as more and more numbers of other colors were added to the display (the lower curve in Fig. 3), also proved to be statistically reliable. The data involved in this demonstration consist of the search times for each individual $S$, summed over target color and display background as before, for displays of one, two, three, four, and five colors, with display densities of 20, 40, 60, 80, and 100 items, respectively, under conditions where the target color was known in advance. As in the previous analysis, only half of the available data for the two-colored, 40-item displays and the three-colored, 60-item displays was used. This represents a factorial combination of 12 $Ss$ by 5 display densities (each representing a different number of displayed colors), for a total of 60 measures. Analysis of variance confirmed a reliable difference among the 5 display types at $P < .01$ ($F = 4.37$; $df = 4/44$). This confirms the conclusion of Green and Anderson, based on a similar analysis, that wrong-colored items can be almost completely ignored in visual search—almost, but not quite.

## SUMMARY

Twelve $Ss$ each viewed a series of 300 displays, which varied in display density, in number of colors used, in the particular color of the target, with either a white or black background, under conditions where $S$ either knew the color of the target in advance, or did not.

Neither the particular color of the target nor the display background had any significant effect on search time. Search time increased regularly with increasing display density. For multicolored displays, when the color of the target was known in advance, search times were considerably shorter than when the target color was unknown. When the color of the target was unknown, search times were not significantly different than those for single-colored displays.

## REFERENCES

CHAPANIS, A., & HALSEY, R. M.   Absolute judgments of spectrum colors. *J. Psychol.*, 1956, **42**, 99–103.

CHRISTNER, C. A., & RAY, H. W.   An evaluation of the effect of selected combinations of target and background coding on map-reading performance: Experiment V. *Hum. Factors*, 1961, **3**, 131–146.

CONOVER, D. W.   The amount of information in the absolute judgment of Munsell hues. *USAF WADC tech. Note*, 1959, No. 58–262.

CONOVER, D. W., & KRAFT, C. L.   The use of color in coding displays. *USAF WADC tech. Rep.*, 1958, No. 55–471.

EDWARDS, A. L.   *Experimental design in psychological research.*   New York: Rinehart, 1950.

ERIKSEN, C. W., & HAKE, H. W.   Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, **50**, 153–160.

GREEN, B. F., & ANDERSON, L. K.   Color coding in a visual search task. *J. exp. Psychol.*, 1956, **51**, 19–24.

GREEN, B. F., McGILL, W. J., & JENKINS, H. M.   The time required to search for numbers on large visual displays. Technical Report No. 36, 1953, Massachusetts Institute of Technology, Lincoln Laboratory.

HALSEY, R. M., & CHAPANIS, A.   On the number of absolutely identifiable hues. *J. Opt. Soc. Amer.*, 1951, **41**, 1057–1058.

HITT, W. D.   An evaluation of five different abstract coding methods: Experiment IV. *Hum. Factors*, 1961, **3**, 120–130.

# RESISTANCE TO EXTINCTION WHEN PARTIAL REINFORCEMENT IS FOLLOWED BY REGULAR REINFORCEMENT

## HERBERT M. JENKINS

*Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey*

The discrimination hypothesis (Humphreys, 1940; Mowrer & Jones, 1945) ascribes the partial reinforcement effect (PRE) to the relative ease of discriminating an abrupt transition to extinction following regular reinforcement as compared to the more gradual transition following partial reinforcement.

If the recent conditions of reinforcement provide the basis of the discrimination, then the PRE should be decreased by the interpolation of regular reinforcement prior to extinction since this would make the transition between recent training and extinction abrupt. This was examined for the case of the free operant response by Keller (1940), by Likely (1958), and by Quatermain and Vaughan (1961). Although in every case the results were that the interpolation of regular reinforcement failed to reduce resistance to extinction, no firm conclusions can be drawn since none of these experiments demonstrated a clear PRE for partial versus regular reinforcement taken separately. Theios (1962) remedied this deficiency in a runway experiment. The PRE was not significantly altered by the interpolation of 25 regular reinforcements, and it was still clearly present, although significantly reduced, after 70 regular reinforcements.

The present experiments use the pigeon's key peck response with a discrete trial procedure. They were carried out independently of Theios' work and afford some of the same comparisons. In Exp. I, the location of a period of partial reinforcement within a longer regime of training under regular reinforcement was varied so that the effect of the amount of regular reinforcement which occurred between partial reinforcement and extinction could be observed where the total amount of training was the same. A baseline was provided by a control group which received all regular reinforcement. The results suggested that the interpolation of regular reinforcement prior to extinction may *increase* resistance to extinction over the level obtained when extinction occurs directly following partial reinforcement. Experiment II was designed to substantiate this finding, to further define the conditions under which it occurs, and to remove certain ambiguities in the interpretation of Exp. I.

## METHOD

*Subjects.*—The Ss were 5-6 yr. old, male White Carneaux pigeons without prior experimental history. They were maintained by restricted feeding at 80% of their free-feeding body weight.

*Apparatus.*—An automatic key pecking apparatus of Skinner design was used. Details of the apparatus have been reported previously (Jenkins, 1961). Reinforcement was a 4-sec. period of access to a tray of mixed grain signaled by lighting the opening to the tray.

*Trials.*—A trial was begun by lighting the translucent plastic response key. The trial was terminated (key light turned off) by a single response, or by external control at the end of 5 sec., whichever occurred first. On reinforced trials, the tray operation followed the response without delay. During the

interval between trials the key light was off, but $S$'s compartment remained illuminated. The time between onsets of successive trials was equally often 15, 30, or 45 sec. in an irregular order. A response made between trials prevented the onset of the next trial for 60 sec. A second response occasionally occurred closely following the first response to a trial. On trials in which the first response initiated an operation of the grain tray, the second response released the tray and thus prevented reinforcement. The second response always placed the 60-sec. delay into effect.

Two measures were recorded on each trial: (a) the occurrence or nonoccurrence of a response within the 5-sec. trial period, and (b) the latency of response in tenths of a second. No latency was entered for trials on which $S$ failed to respond. Intertrial responses (ITRs) were also recorded.

*Shaping of response.*—The $Ss$ were trained to strike the lighted key by reinforcing successive approximations. When several automatically reinforced key responses had occurred, intertrial intervals were introduced. A discrimination between the lighted and unlighted key was formed immediately so that responses were confined almost entirely to the trials from the beginning of training. Two sessions were, with a few exceptions, sufficient to prepare $Ss$ for the automatically programed training.

*Sessions.*—The experiments involved preliminary training, training, and extinction. Preliminary training consisted of four sessions (Exp. I) or five sessions (Exp. II) of regular reinforcement. In each session, 40 reinforced trials were programed. The period of access to the grain tray was reduced from 5 to 4 sec. in the course of these sessions.

Training sessions were under regular or partial reinforcement. In sessions of regular reinforcement 40 reinforced trials were programed as in preliminary training. Sessions of partial reinforcement consisted of 40 reinforced trials plus a number of nonreinforced trials programed in a random sequence subject to the constraint that each half of the sequence contained half of the reinforced trials (20) and half of the nonreinforced trials.

Extinction sessions were run on consecutive days and each consisted of 40 nonreinforced trials. By prior decision, all comparisons among groups were based on the first 10 extinction sessions. Additional sessions were run in the case of most groups in order to more nearly complete extinction.

*Treatment groups, Exp. I.*—The four groups of Exp. I each received 13 sessions of

training beyond preliminary training. Group 13R received 13 sessions of regular reinforcement. The remaining three groups received partial reinforcement for a block of 3 consecutive sessions located in different positions within the sequence of training sessions. These groups are designated 3P-10R (3 sessions of partial reinforcement followed by 10 sessions of regular reinforcement), 9R-3P-1R, and 10R-3P. Forty nonreinforced trials were programed in the first session of partial reinforcement, 60 in the second, and 80 in the third.

Five $Ss$ were assigned to each group so as to match the groups as well as possible on the basis of the means and variance of the latency of response during preliminary training.

*Treatment groups, Exp. II.*—Three groups received different numbers of training sessions under partial reinforcement: Groups 8P $(N = 9)$, 20P $(N = 7)$, and 32P $(N = 7)$. Two groups received different amounts of training under regular reinforcement: Groups 8R $(N = 6)$ and 20R $(N = 6)$. Finally, two groups received different amounts of training under partial reinforcement prior to 12 sessions of regular reinforcement: Groups 8P-12R $(N = 7)$ and 20P-12R $(N = 6)$. The reasons for selecting these conditions will become clearer after the results of Exp. I are reported, but certain features may be noted here. Partial reinforcement was begun at the same point in training (directly following preliminary training), rather than at different points as in Exp. I. The amount of training was extended in order to obtain a clear separation in resistance to extinction between groups given partial reinforcement as against those given regular reinforcement. The design also allows comparison of the effect on extinction of adding, after different amounts of training under partial reinforcement, either more partial reinforcement, or an equal number of sessions of regular reinforcement.

The groups in Exp. II were run in the following order: Groups 8P and 8R concurrently, then Groups 20P, 20R, and 8P-12R concurrently, then Groups 32P and 20P-12R concurrently. Groups run at the same time were matched on the basis of performance in preliminary training as in Exp. I.

In each of the first two sessions of partial reinforcement, 40 nonreinforced trials were programed. Thereafter, each session of partial reinforcement contained 80 nonreinforced trials.

*Additional sessions of extinction.*—All groups in Exp. I and Groups 20R, 20P, and 8P-12R of Exp. II received 2 additional sessions of extinction. Groups 20P-12R and

32P received 4 additional sessions of extinction. In Group 32P 1 S was given 5 and another was given 11 sessions beyond Session 10. With the additional sessions of extinction, there was at least 1 session for every S in which 10 or fewer responses were made. The effect of including responses made in sessions beyond Session 10 on the relative standing of the groups as determined by the first 10 sessions was examined and is reported below.

## RESULTS

*Responses in training.*—The proportion of trials responded to was maintained at very high levels throughout all phases of training. The mean proportion based on both experiments was .976 for preliminary training and .985 for training. There were no systematic differences in this measure among groups or between the periods of partial reinforcement and regular reinforcement within groups.

*Shape of extinction curves.*—The discrete trial method of the present experiments yields extinction curves which typically show a sharp drop from a high to a low level of responding. Individual extinction curves are shown for Ss in Groups 20R and 20P in Fig. 1. Although the average number of responses in extinction was about three times greater in Group 20P than in Group 20R, it appears from Fig. 1 that the curves differ principally in the points at which failures to respond begin, rather than in the abruptness of the decline thereafter.

In order to examine the abruptness of extinction more systematically, two criterion sessions designated A and B were located for each S. Session A was the last one in which 30 or more responses were made and Session B was the first one in which 10 or fewer were made. Figure 2 plots for each group in Exp. II the mean locations (session numbers) of the
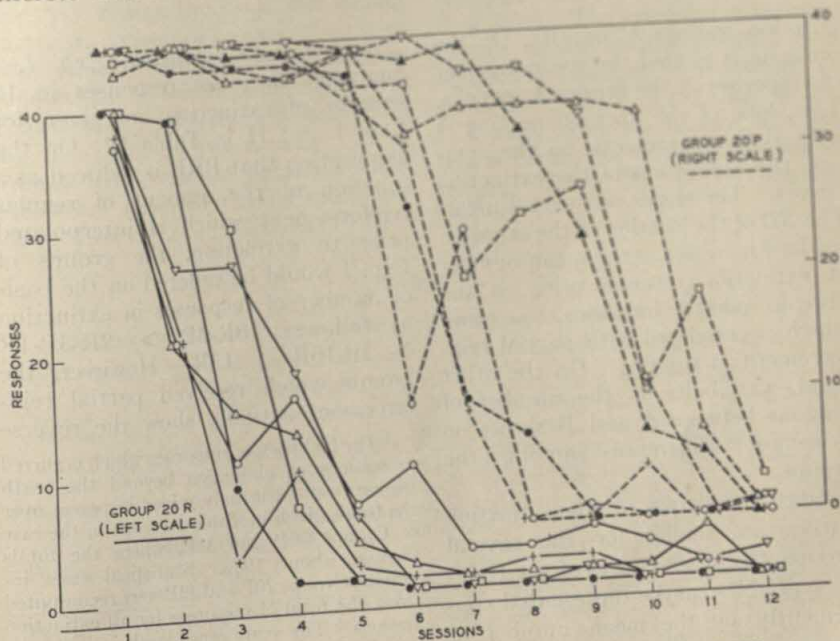


FIG. 1. Responses in extinction for individual Ss of Groups 20P and 20R.
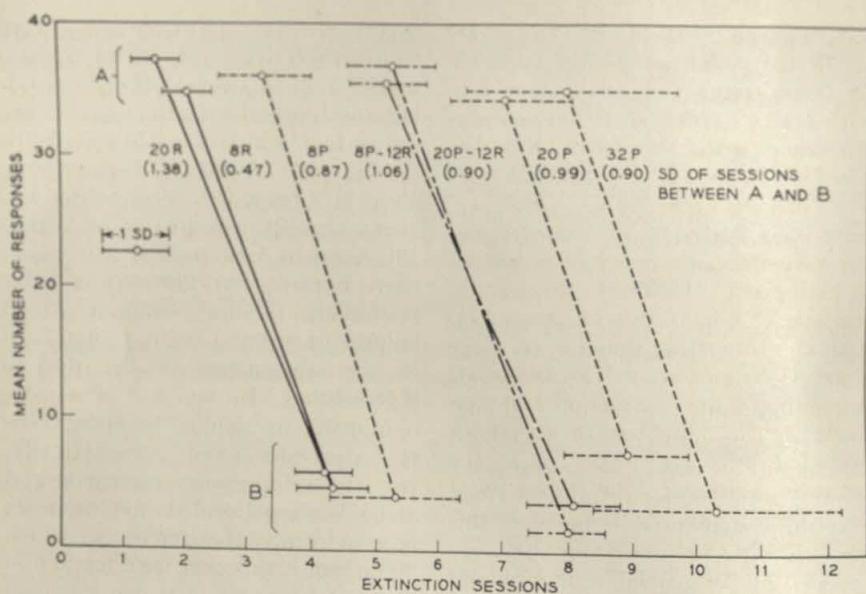
FIG. 2. The average slopes of extinction curves for the groups of Exp. II. (The criterion sessions A and B are explained in the text. The bars represent one *SD* of the location of Sessions A and B, or one-half *SD* to either side of the mean. The numbers in parentheses are one *SD* of the number of sessions between Sessions A and B.)

criterion sessions A and B. On the ordinate is plotted the mean number of responses in the criterion sessions. The slope of the lines connecting A with B thus represents, on the average, the abruptness of the extinction curves. The slopes are very similar. The *SD* of the location of the criterion sessions increases as the mean number of extinction sessions prior to the "break point" increases, a trend which is correlated with partial reinforcement in training. On the other hand, variability in the number of sessions between A and B shows no systematic differences among the groups.

Since the shapes of the extinction curves are similar for the several groups, the major effects of the training variables can be represented conveniently by the mean number of responses in extinction.

*Responses in extinction, Exp. I.—* Summary data for responses in 10 sessions of extinction are given for Exp. I and II in Table 1.[1] On the assumption that PRE is reduced as a function of the amount of regular reinforcement which is interpolated prior to extinction, the groups of Exp. I would be ordered on the basis of number of responses in extinction as follows: 10R-3P > 9R-3P-1R > 3P-10R ≥ 13R. However, the groups which received partial reinforcement actually show the reverse

[1] The inclusion of responses which occurred in sessions of extinction beyond the tenth session resulted in only minor increases over the totals given in Table 2 except in the case of Groups 20P and 32P, where the totals increase about 10%. Statistical tests involving Groups 20P and 32P were recomputed using the total of responses for all extinction sessions. The significance levels in each case remained unchanged from those obtained on the basis of the first 10 sessions of extinction.

ordering: i.e., 10R-3P < 9R-3P-1R < 3P-10R. An analysis of variance for these groups yielded an $F = 3.40$ with $2/12$ $df$, and $P < .10$. The mean difference between the extreme groups was 78.4 responses ($t = 2.60$; $df = 8$, $P < .05$). Group 13R, which received no partial reinforcement had the lowest mean number of responses in extinction. However, meaningful statistical comparisons of this group with the others are difficult to make since 1 $S$ in Group 13R resumed responding late in extinction and thus made a total number of responses exceeded by only 2 other $S$s in the entire experiment. The remaining 4 $S$s in Group 13R had the four lowest totals in the experiment.

The following hypothesis was formulated from the outcome of Exp. I and tested in Exp. II: The interpolation of regular reinforcement between partial reinforcement and extinction does not reduce the PRE, but in fact increases resistance to extinction over that which obtains when extinction follows partial reinforcement directly. Experiment I falls short of establishing this point on two counts. First, a clear PRE was not obtained when extinction followed partial reinforcement directly (i.e., Groups 10R-3P

and 13R were not well separated). In order to obtain a clear PRE against which to evaluate the effect of interpolated regular reinforcement, the amount of training under regular or partial reinforcement was extended in Exp. II. Second, although the hypothesis ascribes the greater resistance to extinction in Group 3P-10R compared to Group 10R-3P to the regular reinforcement which followed partial reinforcement, the difference between these groups could also be ascribed to differences in the amount of regular reinforcement which preceded the partial reinforcement. In Exp. II, partial reinforcement was introduced at the same point in training in order to avoid this ambiguity.

*Responses in extinction, Exp. II.*— A comparison of Groups 8P, 20P, and 32P of Exp. II shows that resistance to extinction was a function of the amount of training under partial reinforcement ($F = 12.18$; $df = 2/12$, $P < .001$). The extension of training from 8 to 20 sessions resulted in a significant increase in resistance to extinction ($t = 4.16$; $df = 14$, $P < .001$), but no further increase resulted from the further extension of training from 20 to 32 sessions ($t < 1$).

TABLE 1

MEANS AND $SD$s OF RESPONSES IN 10 SESSIONS OF EXTINCTION: EXP. I AND II

| Type of Training | Experiment I | | | Experiment II | | |
|---|---|---|---|---|---|---|
| | Group | Mean | SD | Group | Mean | SD |
| Regular | 13R | 100.0 | 71.6 | 8R | 114.5 | 33.2 |
| | | | | 20R | 103.7 | 15.1 |
| Partial | 10R-3P | 118.4 | 48.2 | 8P | 164.3 | 63.9 |
| | | | | 20P | 294.9 | 54.9 |
| | | | | 32P | 298.7 | 54.9 |
| Partial then Regular | 9R-3P-1R | 162.2 | 41.2 | 8P-12R | 247.4 | 34.7 |
| | 3P-10R | 196.8 | 37.8 | 20P-12R | 250.2 | 48.3 |

A comparison of Groups 8R and 20R shows, on the other hand, that when training was under all regular reinforcement, resistance to extinction was unchanged by the extension of training from 8 to 20 sessions ($t < 1$).

A clear PRE emerged only after extended training. Groups 8R and 8P did not differ significantly ($t = 1.64$; $df = 13$, $P < .20$) while Groups 20R and 20P were clearly different ($t = 7.60$; $df = 11$, $P < .001$).

Consider next Group 8P-12R which was switched from partial to regular reinforcement prior to extinction. Resistance to extinction for this group was significantly greater than for Group 8P ($t = 2.91$; $df = 14$, $P < .02$) showing that the addition of regular reinforcement increased resistance to extinction. The increase obtained from adding 12 sessions of regular reinforcement to 8 sessions of partial reinforcement was less, but not significantly less, than the increase resulting from the addition of 12 sessions of partial reinforcement, i.e., the means for the number of responses in extinction in Groups 8R-12P and 20P could not be reliably distinguished ($t = 1.79$; $df = 12$, $P \cong .10$).

When 20 sessions of partial reinforcement were given, the addition of 12 sessions of regular reinforcement (Group 20P-12R) produced no further increase in resistance to extinction. The mean number of responses to extinction was in fact less, although not significantly less, in Group 20P-12R than in either Groups 20P or 32P (20P-12R vs. 20P: $t = 1.42$, $df = 11$, $P < .20$; 20P-12R vs. 32P: $t = 1.55$, $df = 11$, $P < .20$).

The results of Exp. I and II are brought together in Fig. 3 which plots the mean number of responses in extinction as a function of the amount of training under regular reinforcement, the amount of training under partial reinforcement, and for the mixed conditions in which partial reinforcement is followed by regular reinforcement. Experiment I has been coordinated to this plot by ignoring differences in the amount of regular reinforcement which preceded the introduction of partial reinforcement. When treated in this way, the outcome of Exp. I falls into line quite well with that of Exp. II, so that it seems reasonable to believe that variations in the amount of training under regular reinforcement preceding the introduction of partial reinforcement made little difference.

The independent variables in Fig. 3 are the number of training sessions of regular or partial reinforcement *beyond* the sessions of preliminary training. Thus, the dotted lines extrapolated to zero sessions of training represent a guess that preliminary training alone would produce a resistance to extinction equal to that obtained after eight additional sessions of regular reinforcement.

*Relation of latency in training to responses in extinction.*—The amount and type of training had only minor effects on the mean latency of response for groups. There were in fact no significant differences in the mean latencies of response taken over the last three sessions of training in either experiment. The overall mean latency for these sessions was 1.4 sec. However, it was possible to detect in Exp. II a small effect of partial reinforcement when compared to regular reinforcement for the first eight training sessions. The mean latency for partially reinforced Ss ($N = 36$) was .2 sec. longer than for regularly reinforced Ss ($N = 12$). A test of this difference using a covariance analysis with latency during preliminary training as the predictor yielded $t = 2.34$; $df = 45$, $P < .05$. The switch from
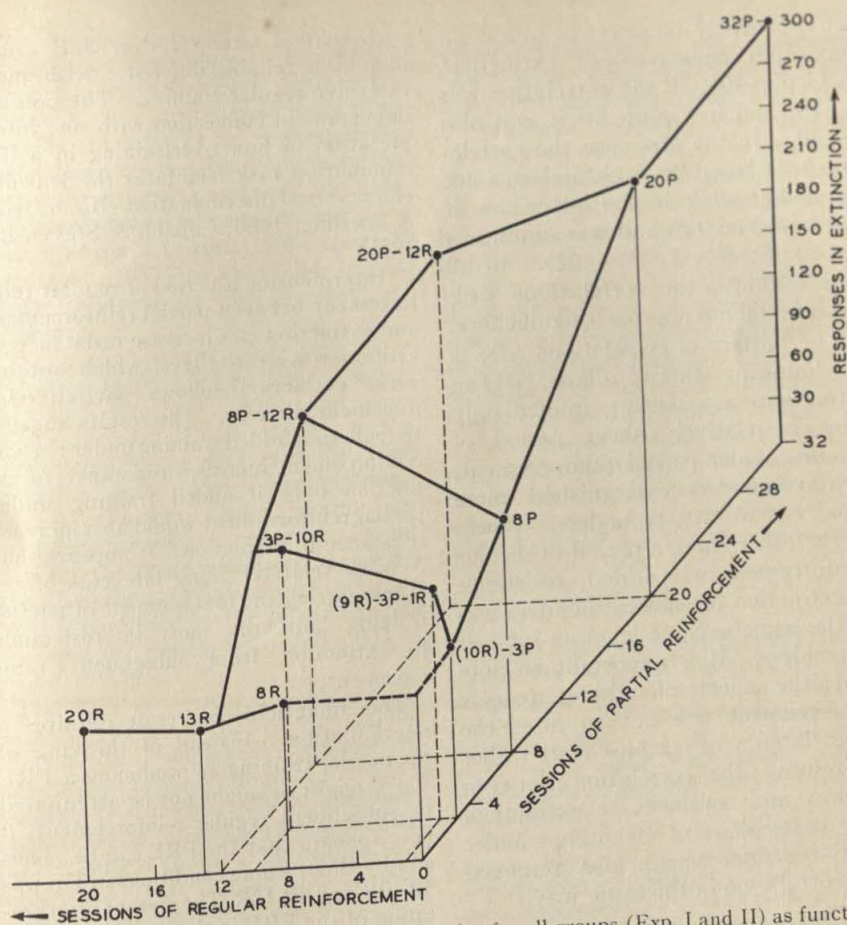
FIG. 3. Mean number of responses in extinction for all groups (Exp. I and II) as function of the amount and type of training beyond preliminary training. (Sessions of regular training preceding partial training in Exp. I have been disregarded and are shown in parentheses [see text].)

partial to regular reinforcement caused no significant change in latency in either experiment.

From the overall results it is apparent that the effects of the amount and type of training on resistance to extinction were not paralleled by effects on group mean latencies of response in training.

The correlation between latencies for individuals within groups and responses to extinction is more interesting since it provides a clue as to

how the addition of training under regular reinforcement after a limited amount of training under partial reinforcement operates to increase resistance to extinction. A substantial negative correlation (longer latencies associated with fewer responses in extinction) was found in the case of two groups: 10R-3P (Exp. I) and 8P (Exp. II). These were the only groups which went directly to extinction after a training period under partial reinforcement that was

not sufficiently extensive to produce a maximum resistance to extinction. In Group 10R-3P the correlation was − .79, and in Group 8P it was also − .79. In the first case the correlation was based on 5 $S$s and was not significant while in the latter case it was based on 9 $S$s and was significant ($t = 3.35$; $df = 7$, $P < .02$). In all other groups, the correlations were low and did not approach significance.

The pattern of correlations reflects the following state of affairs. When extinction was begun immediately after a relatively short period of training under partial reinforcement, slow responders extinguished more rapidly than fast responders. When, on the other hand, a period of regular reinforcement was added, resistance to extinction for slow responders was at the same average level as for fast responders. It is important to note that the effect of adding regular reinforcement was not to alter the mean latency of response, but rather to remove the correlation between latency and resistance to extinction. The extension of training under partial reinforcement also removed the correlation in the same way.

## DISCUSSION

The abruptness of the transition from training to extinction is obviously not the critical functional difference between regular and partial reinforcement. When extensive training is first given under partial reinforcement, its effect persists through a period of training under regular reinforcement involving 480 reinforcements.

Although 480 reinforcements is a large number, it is still a small fraction of the 2,320 training trials under partial reinforcement which were used in Group 20P in order to obtain a clear PRE. It therefore seems possible that the insignificantly lower resistance to extinction for Group 20P-12R in comparison

with that of Group 20P or 32P would become a reliable difference with more extensive regular training. The point is important in connection with one interpretation of how overtraining in a discrimination task facilitates the learning of a reversed discrimination (Birch, Ison, & Sperling, 1960; Capaldi & Stevenson, 1957).

Interpolating a period of regular reinforcement between partial reinforcement and extinction can increase resistance to extinction over the level which obtains when extinction follows partial reinforcement directly. The results suggest the rule that added training under regular reinforcement increases resistance to extinction only if added training under partial reinforcement would also increase resistance to extinction. It appears that $S$s with relatively long latencies of response during the first segment of partial training gain the most in resistance to extinction from subsequent reinforcement.

This unanticipated result requires a reformulation. Instead of thinking of the partial training as producing a PRE which might or might not be attenuated by subsequent regular reinforcement, it now appears that the PRE can be *generated* by the training conditions in Groups 3P-10R and 8P-12R. The present interpretation of the PRE in these groups is that it arose from an interaction of the effects of the nonreinforced trials during the partial training with the subsequent reinforcement during regular training. It is clear that the effect depends upon an interaction since without a prior exposure to nonreinforced trials, resistance to extinction was unchanged by the extension of training under regular reinforcement. Amsel's (1958) account of the PRE, which holds that the combination of conditioned frustration due to nonreinforcement and subsequent reinforcement develops a tolerance for nonreinforcement, provides one conjecture as to how this interaction is mediated.[2]

[2] Results on the occurrence of ITRs suggest that frustration was associated with extinction. During training, the overall mean

Other results (Weinstock, 1954, 1958) indicate that some form of interaction between nonreinforcement and reinforcement can occur across long time intervals (at least 24 hr. separated the partial from the regular training in the present experiments), and even when the reinforcements occur consecutively in a block (Jensen & Cotton, 1960; Lauer & Carterette, 1957) as they did in the regular training of the present experiments.

The present results agree with those of Keller (1940), Likely (1958), Quatermain and Vaughan (1961), and more specifically with Theios (1962), in demonstrating the persistence of the PRE through a period of regularly reinforced training. Further, Theios' finding of a small but significant reduction in the PRE for the group which received more extensive regular reinforcement (Group P-70) is paralleled by the present finding of a lessened resistance to extinction in Group 20P-12R.

The fact of increased resistance to extinction for the partial-regular training sequence is a departure, although indications of it can be found in previous results. In particular, Theios obtained a small increase in resistance to extinction (not significant) when 25 regular reinforcements were added to previous partial training. He considered the possibility that it resulted from the regular reinforcement, but rejected it on the grounds that no increase occurred when regular reinforcement was added to previous regular training. However, the present results show that when the effect is obtained, it arises from an *interaction* of partial and regular reinforcement; a possibility which is overlooked in Theios' argument.

Two implications of the present results for theory may be noted. First, the increased resistance to extinction pro-

duced by the partial-regular sequence poses a problem for the dissonance theory of the PRE (Lawrence & Festinger, 1962) since that theory holds that the processes producing the increased resistance to extinction occur primarily on the nonreinforced trial itself. Second, the remarkable degree of independence between the slope of the extinction curves and the total number of responses in extinction indicates that something like a threshold model of extinction is preferable, at least for conditioning procedures of the type used here, to a decremental model in which an initial probability of response is reduced on each trial by an amount proportional to the number of responses remaining. Models of the latter type lead to exponential decays in which the slope of the curve is correlated with the total number of responses in the curve.

## Summary

A food reinforced, key peck response in the pigeon was used with a discrete trial procedure to study resistance to extinction (number of responses in 10 40-trial sessions) following different amounts of training of three types: partial reinforcement, regular reinforcement, and partial followed by regular reinforcement.

Resistance to extinction showed no systematic change as a function of the amount of regularly reinforced training. On the other hand, it increased as a function of the amount of partial training up to 20 sessions and then leveled off at a value which clearly demonstrated the PRE. Resistance to extinction after training under the partial-regular sequence was never significantly lower than for partial alone. In fact, the addition of regular reinforcement increased resistance to extinction when the amount of prior training under partial was not sufficiently extensive to produce a maximum.

From these results and those of previous experiments the conclusion is well established that the abruptness of the local transition between training and extinction is not critically involved in the PRE.

frequency of ITRs was .5 per 100 trials. In extinction it increased in every group yielding an overall mean of 2.2 per 100 trials. The mean frequency of ITRs showed a peak value in the vicinity of the decline of responses in extinction.

## References

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, 55, 102–119.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, 60, 36–40.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, 50, 195–198.

HUMPHREYS, L. G. Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, 27, 71–76.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, 61, 111–121.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, 60, 41–49.

KELLER, F. S. The effect of sequence of continuous and periodic reinforcement upon the "reflex reserve." *J. exp. Psychol.*, 1940, 27, 559–565.

LAUER, D. W., & CARTERETTE, T. S. Changes in response measures over repeated acquisitions and extinctions of a running habit. *J. comp. physiol. Psychol.*, 1957, 50, 334–338.

LAWRENCE, D. H., & FESTINGER, L. *Deterrents and reinforcements.* Stanford: Stanford Univer. Press, 1962.

LIKELY, F. A. Relative resistance to extinction of aperiodic and continuous reinforcement separately and in combination. *J. gen. Psychol.*, 1958, 58, 165–187.

MOWRER, O. H., & JONES, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, 34, 293–311.

QUATERMAIN, D., & VAUGHAN, G. M. Effect of interpolating continuous reinforcement between partial training and extinction. *Psychol. Rep.*, 1961, 8, 235–237.

THEIOS, J. M. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, 64, 1–6.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, 48, 318–322.

WEINSTOCK, S. Acquisition and extinction for a partially reinforced running response at a 24-hr. inter-trial interval. *J. exp. Psychol.*, 1958, 56, 151–158.

# PARTIAL REINFORCEMENT, CONTINUOUS REIN-FORCEMENT, AND REINFORCEMENT SHIFT EFFECTS [1]

STEWART H. HULSE

*Johns Hopkins University*

Hulse (1962) has shown that partial reinforcement (PRF) of running in a straight alley produces sharp discrimination of reward stimuli in the goal box, but continuous reinforcement (CRF) does not. Partially reinforced rats learned to lick from a drinking tube much faster than continuously reinforced rats, and the number of licks they emitted on nonreinforced running trials rapidly decreased to a very low level. Taken as a whole, the data indicated that behavior in the goal box was much more critically focused on the reward and its stimulus properties if PRF as compared with CRF were used.

There is an important implication of these experimental results. During the course of training with PRF, *changes* in the stimulus properties of the reward ought to be sharply discriminated. For example, on the assumptions that sweetness is a valid stimulus continuum for a reinforcing substance (Guttman, 1953) and response rate is positively correlated with sweetness, shifts from low to high concentrations of a saccharin reinforcer ought to produce prompt and direct increases in response rate. Discrimination of a stimulus change of this sort, and appropriate changes in response rate, ought to occur to a much lesser degree, if at all, with CRF. This would be true since CRF

does not appear to bring behavior under the stimulus control of the reinforcer to the extent that PRF does.

The data reported here constitute a test of this implication. In one experiment, rats were conditioned on low and high concentrations of saccharin with PRF and CRF. Then, for half the rats, reinforcement concentrations were switched from low to high or from high to low. In a second experiment, essentially the same procedure was used, except that rats were switched only from low to high concentrations, and more extreme differences in concentration were used. Instrumentally conditioned licking was used as a response. This response is particularly appropriate, since the tongue is intimately involved in the sensory dimension of taste, and since licking rates vary quite consistently with the concentration of a sweet reinforcing stimulus (Hulse & Bacon, 1962).

## EXPERIMENT I

### Method

*Subjects.*—The *S*s were 64 naive male albino rats, 85 days of age, of the Sprague-Dawley strain obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The *S*s had no experience drinking from tubes prior to their use in the experiment.

*Apparatus and procedure.*—The apparatus and the general procedure for treating licking as an instrumental response have been described in detail elsewhere (Hulse, 1960; Hulse, Snyder, & Bacon, 1960). In brief, the apparatus consisted of a $6.5 \times 10 \times 7$ in. wooden box which was suitably lighted, ventilated, and sound-shielded. A piece of $\frac{1}{4}$-in. Plexiglas was located in front of a hole

in one wall of the box. The Plexiglas had a 1 × 2 cm. vertical slot cut into it. The S had access through the slot to a drinking tube.

Each lick on the drinking tube operated an electronic relay which operated, in turn, a programing circuit. Liquid reinforcements of a specified volume were delivered to S through the tube by means of an infusion pump operated by the programmer. For partial schedules of reinforcement, the programmer also operated a solenoid-driven lever. The lever slightly squeezed the pressure tubing leading from the pump to the drinking tube. After the reinforcement was delivered, the tubing returned to its normal shape and produced a slight negative pressure in the fluid system. The negative pressure drew the fluid column about 1 mm. inside the tip of the drinking tube. This assured that S could not make contact with the fluid on nonreinforced licks.

Two drinking tubes were used in the experiment. During the first 4 days of training, a brass tube with a 2-mm. fluid hole was used. On Day 1, this tube projected through the slot in the experimental box; on Days 2 to 4, the tube was gradually withdrawn through the slot. On Day 5, a plastic nipple with a 2-mm. fluid hole was introduced behind the slot and used for the rest of the experiment. This tube had a $\frac{1}{16}$-in. diameter brass electrical contact located just below the fluid hole.

The experimental design that was used during initial training incorporated two schedules of reinforcement (CRF and FR8), and two concentrations of reinforcement (1 gm. and 10 gm. of saccharin added to 1 l. of tap water). Sixteen Ss were randomly assigned to each of the four groups called for by this design. After initial training, the concentrations of reinforcement were switched for half the Ss trained with each schedule of reinforcement. For both the CRF and FR8 schedules of reinforcement, the postswitch design thus included one group switched from the low to the high concentration, one group switched from high to low, and two control groups which continued on their initial concentrations ($N = 8$ in each group).

Upon receipt from the breeder, Ss were put in colony cages and maintained on ad lib. Purina lab chow and water (available from cans) for 10 days. Ten days of taming followed. On Day 1 of taming, Ss were transferred to individual cages and placed on a daily deprivation diet of 10 gm. ground Purina lab chow mixed with 20 cc of tap water. They stayed on this deprivation schedule throughout the remainder of the experiment. Also, on each taming day, Ss

were taken in groups of approximately 15, freely handled, and allowed to explore in a large wooden box for 30 min. They received their daily ration of wet mash in their home cages after this procedure was completed.

Following taming, all Ss received 22 days of initial conditioning of the instrumental licking response. On Day 1, each S was placed in the apparatus and permitted 300 licks from the brass tube on CRF. The concentration of reinforcement variable was introduced immediately. The size of each reinforcement was .005 cc of fluid. The plastic nipple was introduced on Day 5. The number of licks permitted all Ss and the ratio of reinforcement for the partial Ss gradually increased so that by Day 7, all Ss made 1,000 licks, and by Day 18, the partial Ss were on their final FR8. This regime was continued through Day 22.

On Day 23, the concentration of saccharin for Ss in the switched groups was shifted from high to low or from low to high, depending upon the concentration that was used during initial training. This regime was continued from Day 23 through Day 29. On Day 30, Ss in the switched groups were shifted back to their original concentrations and run under this regime through Day 36. The Ss in the control groups remained on their initial concentrations throughout the 36 days of training.

On Day 37, 3 days of extinction began. Each day, Ss were placed in the apparatus and allowed to lick the dry tube for 3 min. All conditions remained the same as those prevailing during training, except that the pump and solenoid-driven lever were disconnected from the programmer.

The Ss were fed their daily ration of wet mash 20 to 30 min. after they had been run. The amount of water added to the mash was adjusted to account for the fluid obtained in the apparatus.

During training, the total time that each S required to emit its 1,000-lick allotment was recorded for each day. These times were transformed to reciprocals and multiplied by 1,000 to give a measure of average licking rate in licks per second. During extinction, the total number of licks that each S emitted during the daily 3-min. extinction period was recorded.

## Results

*Preswitch performance.*—By Day 22, the last training day before the first switch in saccharin concentration,

CRF produced faster licking rates than PRF, and the high concentration of saccharin produced faster licking rates than the low concentration of saccharin. An analysis of variance for Day 22 for Ratio and Concentration of Reinforcement yielded an $F$ for Ratio of 15.83 ($df = 1/60$, $P < .01$) and an $F$ for Concentration of 59.11 ($df = 1/60$, $P < .01$). The interaction between the two variables was not significant ($P > .05$).

*First switch.*—The effect of the first switch in concentration, which occurred on Day 23, was quite different for CRF and PRF. Figure 1 shows that with PRF an orderly and appropriate change in rate occurred from Day 22 to Day 23. If the switch was from a high to a low concentration, licking rate decreased. If the switch was from low to high, licking rate increased. With CRF, on the other hand (Fig. 2), licking rates decreased regardless of the direction of the switch in concentration.

Statistical analyses of these results were carried out in two ways. First, analyses of variance were used on means for the groups that were switched in concentration (Table 1). These analyses were done separately



FIG. 2. Effect of shifts in concentration of reinforcement for CRF in Exp. I. (Shifts occurred between Days 22 and 23, and between Days 29 and 30.)

for the CRF and PRF conditions, and they included pre- and postswitch days as a variable. Second, analyses of variance were used on means for both the control and switched groups (Tables 2 and 3). These were done separately for both the CRF and PRF conditions and for Days 22 and 23.

Table 1 shows a highly significant Treatments × Days interaction for the PRF condition. By $t$ test, the licking rate of the PRF High-Switched group decreases from Day 22 to Day



FIG. 1. Effect of shifts in concentration of reinforcement for PRF in Exp. I. (Shifts occurred between Days 22 and 23, and between Days 29 and 30.)

TABLE 1

ANALYSES OF VARIANCE OF SWITCHED-GROUP LICKING RATES DURING TRAINING

| Source | df | MS | | | |
|---|---|---|---|---|---|
| | | PRF | | CRF | |
| | | Days 22–23 | Days 29–30 | Days 22–23 | Days 29–30 |
| $Ss$ | 15 | | | | |
| Treatments (T) | 1 | 0.98 | 0.21 | 1.67 | 0.84 |
| Error | 14 | 1.09 | 0.70 | 1.35 | 0.75 |
| Days (D) | 1 | 1.62 | 0.01 | 9.57** | 0.05 |
| $Ss × D$ | 15 | 0.67 | 0.30 | 0.56 | 0.26 |
| T × D | 1 | 7.22** | 2.64** | 1.32 | 2.65** |
| Error | 14 | 0.20 | 0.13 | 0.51 | 0.09 |

** $P < .01$.

## TABLE 2

ANALYSES OF VARIANCE OF LICKING RATES DURING TRAINING FOR PRF GROUPS

| Source | df | MS | | | | |
|---|---|---|---|---|---|---|
| | | Day 22 | Day 23 | Day 29 | Day 30 | Day 36 |
| Concentration (C) | 1 | 12.50** | 7.70** | 6.57** | 7.22** | 5.36** |
| Switch (Sw) | 1 | 0.01 | 0.38 | 0.94 | 0.67 | 3.06* |
| C × Sw | 1 | 0.01 | 1.17 | 1.95* | 0.36 | 0.43 |
| Error | 28 | 0.43 | 0.54 | 0.34 | 0.42 | 0.62 |

\* $P < .05$.
\*\* $P < .01$.

23 ($t = 6.36$, $df = 14$, $P < .01$), and the licking rate of the PRF Low-Switched group increases from Day 22 to Day 23 ($t = 2.27$, $df = 14$, $P < .05$). For the CRF condition, Table 1 shows that only Days is a significant variable. Licking rates on Day 23 are much lower than on Day 22, and this is true regardless of the direction of the switch in concentration. Tables 2 and 3 show that for both PRF and CRF on Day 22 only the Concentration variable is significant. High concentrations produced faster licking rates than low concentrations. On Day 23, however, the pattern of significant effects is quite different for PRF and CRF. For PRF, the Concentration variable once again produces the only significant effect. Licking rate is correlated with the

## TABLE 3

ANALYSES OF VARIANCE OF LICKING RATES DURING TRAINING FOR CRF GROUPS

| Source | df | MS | | | | |
|---|---|---|---|---|---|---|
| | | Day 22 | Day 23 | Day 29 | Day 30 | Day 36 |
| Concentration (C) | 1 | 8.92** | 0.92 | 1.62 | 4.97** | 11.56** |
| Switch (Sw) | 1 | 0.16 | 7.22** | 0.10 | 0.13 | 0.12 |
| C × Sw | 1 | 0.31 | 1.20 | 0.32 | 0.09 | 0.02 |
| Error | 28 | 0.32 | 0.90 | 0.45 | 0.34 | 0.27 |

\*\* $P < .01$.

concentration used on Day 23 for both switched and control groups. For CRF, only the Switch variable is significant. Licking rates are generally lower for the switched groups than they are for the control groups. By $t$ tests, both the High-Switched and the Low-Switched groups differ from the High-Control group ($t$'s = 2.72 and 2.83, $df = 28$, $Ps < .05$ and .01, respectively), but neither switched group differs significantly from the Low-Control group $Ps > .05$).

Terminal performance on the concentrations of the first switch was quite different for PRF and CRF.[2] For PRF on Day 29 (Fig. 1), licking rates for the Low-Switched group reached about the same level as those of the High-Control group. Licking rates for the High-Switched group, however, showed an increase following Day 23 and reached an asymptote that was much higher than that of the Low-Control group. Table 2 shows significant effects for both Concentration and the interaction between Concentration and Switch in Concentration. By $t$ test, the Low-Switched group does not differ from the High-Control group ($P > .05$), but the High-Switched group licks significantly faster than the Low-Control group ($t = 2.90$, $df = 28$, $P < .01$). The results for CRF (Fig. 2) show a pattern of differences among the groups which is similar to that for PRF, but the magnitudes of the differences are much smaller. Table 3 shows that for CRF none of the experimental variables produced significant effects on Day 29.

[2] Analyses of variance for the last 3 days of the first and second switches revealed no significant effects due to Days or the interaction of Days with any of the other variables. The data thus provide no evidence that the groups had not reached asymptotic performance by Day 29 or by Day 36.

*Second switch.*—With one major exception, the effect of the second switch in concentration on Day 30 was quite similar for both PRF and CRF. Figures 1 and 2 show that the licking rates of the switched groups changed in the appropriate direction when these groups returned to their original training concentrations. This was true regardless of ratio of reinforcement. The extent of the change in rate, however, was quite different for PRF and CRF. With CRF, Fig. 2 shows that the switched groups reached licking rates which were essentially the same as the control groups. With PRF, however, Fig. 1 shows that the Low-Switched group changed to a performance level which was considerably above that of the Low-Control group. By Day 36, the PRF High-Switched group also responded at a faster rate than the High-Control group.

Table 1 shows a significant Treatments × Days interaction for the switched groups for Days 29 and 30 for both PRF and CRF. Tables 2 and 3 show a significant Concentration effect on Day 30 for both PRF and CRF, but none of the other variables are significant. On Day 36, however, there is a significant Switch effect for PRF in addition to significant Concentration effects for both PRF and CRF. Although both PRF switched groups performed above their controls on Day 36, the difference between the Low-Switched and the Low-Control group is the only one that reaches significance ($t = 2.16$, $df = 28$, $P < .05$).

*Extinction.*—Figure 3 shows that, in general, the PRF groups emitted more licks on the dry drinking tube than the CRF groups. Moreover, with PRF, the switched groups were far more resistant to extinction than the nonswitched groups. With CRF,
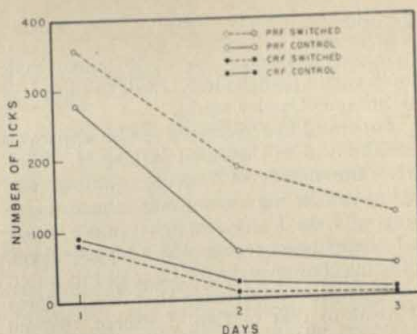


FIG. 3. Resistance to extinction in Exp. I for PRF and CRF Switched and Control groups.

however, the data suggest that the switched groups were less resistant to extinction than the nonswitched groups.

An analysis of variance on mean number of licks for all groups across the 3 days of extinction shows significant effects for Ratio of Reinforcement ($F = 70.48$, $df = 1/56$, $P < .01$), Days of Extinction ($F = 102.93$, $df = 2/126$, $P < .01$), Switch in Concentration ($F = 5.78$, $df = 1/56$, $P < .05$), Ratio × Switch ($F = 9.10$, $df = 1/56$, $P < .01$), and Ratio × Days ($F = 41.46$, $df = 2/112$, $P < .01$). For PRF, the mean difference between the switched and nonswitched groups is significant ($t = 6.61$, $df = 56$, $P < .01$). For CRF, this difference is not significant ($t = 0.76$, $df = 56$, $P > .05$).

## EXPERIMENT II

Experiment II was run to check on the generality of some of the phenomena which resulted from the first switch of Exp. I. Experiment II incorporated more days of training before the switch took place and used more extreme differences in the concentration of the saccharin reinforcement.

## Method

*Subjects.*—The *S*s were 30 naive male albino rats, approximately 75 days of age, of the Sprague-Dawley strain.

*Apparatus and procedure.*—The apparatus was identical to that used in Exp. I. Similarly, the details of housing, taming, and feeding of *S*s were essentially the same as those of Exp. I. A somewhat more severe daily deprivation regime was adopted: 9 gm. of Purina chow mixed with 18 cc of tap water.

The experimental design that was used was identical to that of Exp. I except that no High-Switched groups were run. The post-switch design thus included six groups with an *N* of 5 in each group. Reinforcement concentrations were either .50 gm. or 10 gm. of saccharin added to 1l. of tap water. As before, ratio of reinforcement was either CRF or FR8. The size of each reinforcement was .0043 cc of fluid.

Initial training in Exp. II was identical to that of Exp. I with the following exceptions. Only 600 licks were permitted each day. The PRF *S*s were on their final FR8 schedule by Day 16, and all *S*s continued initial training through Day 32. The PRF *S*s thus received 17 days of training on FR8 instead of the 5 days that the PRF *S*s received in Exp. I. Following initial training, concentration of reinforcement was changed from low to high on Day 33 for the PRF and CRF Switched groups, and the *S*s continued on the new regime through Day 37.

## Results

The immediate effects of the switch from a low to a high concentration are qualitatively exactly the same as



FIG. 5. Effect of shift in concentration of reinforcement for CRF in Exp. II. (The shift occurred between Days 32 and 33.)

those obtained in Exp. I. Figure 4 shows that with PRF, the licking rates of the Switched group increased markedly from Day 32 to Day 33. Comparison of the data shown in Fig. 4 with those shown in Fig. 1 suggests that the increase in rate was considerably greater under the conditions of Exp. II than under those of Exp. I. Figure 5 shows that with CRF, licking rates of the Switched group declined, as they did in Exp. I.

An analysis of variance used on the groups means of Days 32 and 33 for PRF shows the following effects to be significant: Treatments ($F = 14.76$, $df = 2/12$, $P < .01$), Days ($F = 4.66$, $df = 1/14$, $P < .05$), and Treatments $\times$ Days ($F = 7.49$, $df = 2/12$, $P < .01$). The significant Treatments effect occurs primarily because of the large and consistent difference in performance between the High-Control and the Low-Control groups, while the significant Treatments $\times$ Days effect reflects the sharp change in performance of the Switched group from Day 32 to Day 33. An analysis of variance for CRF shows Treatments to be the only significant effect ($F = 12.61$, $df = 2/12$, $P < .01$). A third analysis of variance, which compared the performance of the



FIG. 4. Effect of shift in concentration of reinforcement for PRF in Exp. II. (The shift occurred between Days 32 and 33.)

CRF and PRF Switched groups on Days 32 and 33, shows a significant Days × Ratio of Reinforcement interaction ($F = 9.14$, $df = 1/8$, $P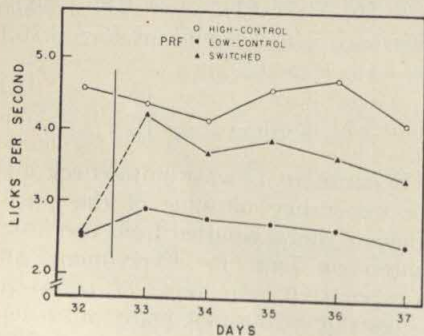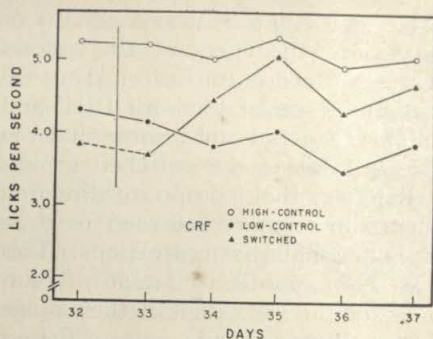 < .05$). This corroborates the results shown by the separate analyses for PRF and CRF regarding the effects of the switch in concentration.

Following the first day of the switch in concentration, the performance of the CRF and PRF Switched groups was quite similar. Both groups reached a licking rate which was between the rates of the High-Control and Low-Control groups. An analysis of variance of means obtained by averaging rates on Days 36 and 37 shows a significant effect due to Treatments ($F = 18.86$, $df = 2/22$, $P < .01$) and to Ratio of Reinforcement ($F = 15.57$, $df = 1/22$, $P < .01$). The interaction between these two variables is not significant. By $t$ test, the Switched groups responded significantly faster than the Low-Control groups ($t = 3.62$, $df = 22$, $P < .01$) and significantly slower than the High-Control groups ($t = 2.50$, $df = 22$, $P < .05$).

## DISCUSSION

The licking behavior that occurs on the day of a sudden change in the concentration of the reinforcer indicates that rats who are trained with PRF are conditioned to respond to the reinforcing stimulus in a way which is quite different from rats who are trained with CRF. With PRF, the sweetness and other stimulus properties of the reinforcer are sharply defined for S. This happens because of the contrast in stimulus conditions which occurs when responses are sometimes followed by a drop of saccharin and sometimes followed by nothing. If the concentration of the saccharin suddenly changes, the change is relatively easy to detect, and response rate immediately increases or decreases in the direction of the change. In effect, behavior comes under critical and orderly control of the stimulus properties of the reinforcer through discrimination training.

A different situation prevails when CRF is used. Here, each successive response is followed by the same sweet stimulus, and there is no stimulus contrast from one response to the next. With CRF, S learns to attack the drinking tube in a way which is vigorous, but relatively speaking, blind. If the concentration of the saccharin reinforcer changes, stimulus generalization decrement prevails, and behavior is markedly disrupted. Pure generalization decrement will hold the first time that S meets a sudden change in concentration, but not, of course, on later occasions. With the first change, S receives new information about reinforcement conditions and is, in fact, trained through contrast to discriminate somewhat more about the stimulus properties of the reinforcer. A second change in concentration comes as less of a surprise to S, and as Exp. I shows, response rates immediately increase or decrease appropriately.

The behavior that appears on the days which follow a change in reinforcement concentration, and the behavior that appears during extinction, indicates that experience with a particular set of reinforcement conditions on PRF produces residual effects which have a long-lasting influence on response strength. The discrimination training that PRF provides apparently yields a *learned* connection between licking behavior and a particular stimulus property of the reinforcer, such as its sweetness.

First, licking rates for the PRF switched groups reach postshift asymptotes which often lie between the licking rates that are characteristic of the high and low concentrations of saccharin. The Ss behave much as if they were averaging the intensities of the pre- and postswitch reinforcing stimuli. Stimulus averaging of this sort seems to be most easily demonstrated when concentrations are changed from high to low (cf. Fig. 1). On the first postswitch day of the second

change in concentration of Exp. I, for example, the response rate of the PRF Low-Switched group falls sharply. But it remains well above the response rate of the Low-Control group. In Exp. II, a compromise in rate is also obtained when concentrations are changed from low to high. The PRF Switched group increases its licking rate on the first postswitch day, but then the rate falls to an asymptote which is again about half way between the rates of the control groups. Sometimes, stimulus averaging and comprises in response rate also appear with CRF. But when they do, they are far less easily obtained than they are with PRF. In Exp. I, there is no indication of a consistent compromise in response rate for CRF. However, in Exp. II, which involved more training days and larger differences in reinforcement concentration, a compromise in response rate is clearly obtained for CRF as well as for PRF. Apparently the more extreme conditions of Exp. II are required, however, before a compromise in rate will appear with CRF and instrumentally conditioned licking. In this connection, Premack and Hillix (1962) trained rats to lick 4% and 16% sucrose solutions from a conventional drinking tube, a procedure analogous to CRF conditioning as defined here, and then shifted the concentration of sucrose for these rats to 32%. The licking rates on the new concentration were consistently lower than the rate of rats maintained on 32% sucrose throughout the experiment. Apparently, experience with a particular reinforcing stimulus on CRF can make that stimulus have long-lasting effects on behavior under some conditions. The important point is that, for the present at least, these conditions appear less well defined for CRF than they do for PRF.

Second, if experience with a particular set of reinforcement conditions on PRF produces residual effects which permanently influence behavior, extinction should provide a sensitive test for these effects. This would be true since extinction removes $S$ from direct exposure

to the primary reinforcing stimulus. The data suggest that the control of behavior by the reinforcing stimulus is indeed permanent and that it does transfer to extinction. First, the extinction data of Exp. I show that a past history of shifts in the concentration of the reinforcer serves to markedly increase resistance to extinction when PRF is used. Further, Hulse (1958), Wagner (1961), and Hulse and Bacon (1962) have shown that with PRF resistance to extinction increases as amount of reinforcement increases. This is true whether amount of reinforcement is defined in terms of weight of food or sweetness of saccharin solutions.

None of these things are true when CRF is used. In Exp. I the CRF groups that had a past history of shifts in concentration of the reinforcer were, if anything, less resistant to extinction than the control groups. In other experiments which have used CRF and varied amount of reinforcement, the extinction results are at best unpredictable. Sometimes resistance to extinction increases as amount of reinforcement increases (Zeaman, 1949), sometimes it decreases (Armus, 1959; Hulse, 1958), and sometimes it does not vary with amount of reinforcement at all (Hulse & Bacon, 1962). Since $S$ receives no discrimination training for the reinforcer with CRF, it is perhaps not surprising that these inconsistent results should be obtained. The data suggest that with CRF, $S$ is stimulus bound. Response strength will be consistently correlated with some stimulus property of the reinforcer as long as $S$ is directly exposed to it, but change the reinforcer—or remove it—and the correlation vanishes.

Finally, none of the data in either experiment reveals the slightest trace of elation, depression, positive contrast, or negative contrast (Pubols, 1960). The only time that a switched group licks either faster or slower than a control group, in a way which might suggest a contrast effect, occurs after the initial change in concentration of Exp. I. Here the rate of the CRF High-Switched

group drops below that of the Low-Control group, but it immediately bounces back to a point which is higher than the Low-Control group. As we have seen, it seems most meaningful in the context of the present experiments to view changes in rate of this sort as due to stimulus generalization decrement. In the same connection, it is interesting to note that most of the significant contrast effects that are reported in the literature are negative contrast effects (Spence, 1956; Pubols, 1960). Since all the earlier experiments have used CRF, perhaps generalization decrement was at least as important as some emotional-motivational factor in determining the contrast effects that were obtained.

## SUMMARY

Instrumentally conditioned licking was studied in two experiments as a function of ratio of reinforcement and shifts in the concentration of a saccharin reinforcer. With partial reinforcement (PRF), a shift in concentration produces an immediate change in licking rate in the direction of the shift. With continuous reinforcement (CRF), the immediate reaction to the shift is always a decrease in response rate. With PRF, asymptotic response rates reached several days after a shift are often compromises between the rates of control groups maintained on high and low concentrations. With CRF, a compromise of this sort appears only when relatively large amounts of training and relatively extreme differences in concentration of reinforcement are used.

The data suggest that, with PRF, behavior is more critically and permanently under the control of reinforcement stimuli than with CRF. This happens because PRF provides discrimination training for reinforcement stimuli, but CRF does not.

## REFERENCES

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as a reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HULSE, S. H. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

HULSE, S. H. Discrimination of the reward in learning with partial and continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 227–233.

HULSE, S. H., & BACON, W. E. Supplementary report: Partial reinforcement and amount of reinforcement as determinants of instrumental licking rates. *J. exp. Psychol.*, 1962, **63**, 214–215.

HULSE, S. H., SNYDER, H. L., & BACON, W. E. Instrumental licking behavior as a function of schedule, volume, and concentration of a saccharin reinforcer. *J. exp. Psychol.*, 1960, **60**, 359–364.

PREMACK, D., & HILLIX, W. A. Evidence for shift effects in the consummatory response. *J. exp. Psychol.*, 1962, **63**, 284–288.

PUBOLS, B. H., JR. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

ZEAMAN, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

# INFERENTIAL BEHAVIOR IN CHILDREN AS A FUNCTION OF AGE AND SUBGOAL CONSTANCY [1]

TRACY S. KENDLER     AND     HOWARD H. KENDLER

*Barnard College*           *New York University*

Problem solution sometimes requires the integration of units of behavior that are already in the repertoire of *S* but have not previously been used in conjunction with one another. When the efficiency of integration increases gradually as a function of the number of trials the process is variously described as trial-and-error, instrumental, or selective learning. When an efficient solution occurs on the first trial, without any preceding trial-and-error behavior, the process is variously designated as reasoning, insight, or inference. The present experiment is one of a series that seeks to explore the relationship between these two processes in children (Kendler & Kendler, 1956; Kendler, Kendler, Pliskoff, & D'Amato, 1958; Kendler & Kendler, 1961). These studies have used an experimental paradigm, derived from Hull (1935, 1952), in which *S* is trained on three discrete behavior segments (A-B, X-Y, and B-G) and then presented with a test situation in which *S* is instructed to get G when only A and X are available. Problem solution requires the assembly of A-B and B-G.

The purposes of the present study were (*a*) to determine whether the ability to infer, as measured by these particular operations, increases with age and (*b*) to analyze the role of the

integrating B stimulus in the inferential process.

## METHOD

*Experimental design.*—There were four experimental groups, 32 *S*s in each, arranged in a 2 × 2 factorial design. The main effects were age, kindergartners (K) vs. third-graders (3rd), and the constancy of training and test subgoals, constant (C) vs. switched (S).

*Subjects.*—The *S*s were 137 children drawn from the Le Conte Public School in Berkeley, California. Nine *S*s were eliminated, 5 of them because they made simultaneous A and X choices, and 4 due to inadvertent errors in experimental procedure. The data to be reported are based on the remaining 128 children, of whom 63 were boys and 65 girls. The kindergartners' mean age was 68.7 mo., range from 61 to 74. The third-graders' mean age was 103.9 mo., range from 96 to 115. Within each age level *S*s were randomly assigned to C and S groups without regard to age or sex.

*Apparatus.*—The portable aluminum apparatus used consisted of three distinct square panels 17.5 cm. on a side. Each panel, which could be exposed to *S*'s view singly or in combination with the others, corresponded to one habit segment. The center panel provided for the B-G segment. On its blue anodized surface was a circular opening 2.4 cm. in diameter into which *S* could drop the objects that served as subgoals in this experiment, namely, a glass marble with a 1.8-cm. diameter and a steel ball bearing with a 1.3-cm. diameter. If *S* dropped the correct subgoal into the circular opening, a small, shiny, gold fairy-tale or nursery-rhyme charm was propelled to a trough near the bottom of the panel. The incorrect subgoal delivered no reward. The charms, e.g., "little Bo Peep," "the cow that jumped over the moon," "the gingerbread man," etc., which were presented in a set sequence, served as the major goal (G).

The two side panels corresponded to the A-B and X-Y segments. The left panel was anodized pink and the right anodized yellow.

Each panel was equipped with a button which, when pressed, closed a circuit that led to the delivery of the appropriate subgoal (glass marble or ball bearing) to a trough near the bottom of that panel.

*Procedure.*—All Ss were run individually and completed in one experimental session. The session began with the administration of the Peabody Picture Vocabulary Test (Dunn, 1959), which took from 5 to 15 min. and was followed immediately by training on the inference apparatus. One of the side panels was opened and S was told, "Press this button and see what happens." When the subgoal was delivered, S was instructed to pick it up, look at it, then return it to E so that he might have another turn. This side panel was closed, the other side panel opened, and the procedure repeated. After S had one forced trial on each panel, the procedure was repeated with the order of the sequences reversed. Thus at this point in training S had made two responses on the A-B and on the X-Y segments in an ABBA order. After these four forced trials, the doors of both side panels were opened and S was shown one of the subgoals and directed to, "Press the button that will get one like this." The procedure was repeated with each of the subgoals presented in an ABBA-BAAB order until the criterion of six successive correct responses was reached.

The next step of the training started with E opening the middle panel (after closing the two side ones) and directing the attention of S to a small window through which he could see a charm. He was told that it was a fairy-tale or nursery-rhyme charm which he would soon get an opportunity to examine closely. The aperture was pointed out and S was informed that if he dropped "the right thing" into that hole, the charm would drop into the tray. Then the marble and the ball bearing were placed in front of the B-G panel and E said, "One of these two little things is the right thing. Take one of them and drop it in the hole to see if it makes the charm come out." On the next trial S was again instructed to drop in the one that would make the charm come out. After that, training proceeded until a criterion of four successive correct responses was reached. Ss who made two or less errors received no additional instructions while those who committed three or more errors were again urged to drop in the one that would make the charm come out. The relative positions of the marble and ball bearing (i.e., left and right) were varied in a random order *after each correct response* throughout the training on the B-G segment.

After preliminary training the test trial was introduced with the following instructions: "Would you like to see another charm? Very well, this time I won't put out any little things, but I will open all the doors. If you do what you are supposed to, you can make the charm come out. Go ahead." The S was allowed 60 sec. in which to make any response that he chose. If he did not press the A or X button during this time, E said, "Which button should you press to help you get the charm? Go ahead." After S pressed either button, he was allowed another 60 sec. to complete the sequence by dropping either subgoal into the B-G aperture before the trial was terminated. Thus all Ss made either an A or X response and the trial was terminated either when S made a major goal response or when 60 sec. had elapsed since the subgoal had appeared.

For half of the children at each age level the subgoals were switched (S) between preliminary training and the test trial so that if S made an A choice, he obtained the Y subgoal while if he made an X choice he obtained the B subgoal. For the remainder the subgoals remained in a constant position (C) during training and testing.

In order to minimize contamination from position and subgoal preferences, they were assigned in counterbalanced order: (a) Half of the Ss in each experimental group had the right panel serve as the A-B segment while the left panel served as the X-Y segment. The opposite arrangement applied to the remaining half. (b) For half of the Ss in each experimental group the marble was B and the ball bearing Y; for the remainder the ball bearing was B and the marble Y.

At the end of the experimental session each S was asked to keep the game a secret until everyone had a turn. It should be emphasized that at no point in the training or test did E describe the subgoal by name.

## RESULTS AND DISCUSSION

*IQ scores.*—The mean PPVT IQ scores for the four experimental groups were as follows: K-C (Kindergarten-Constant), 105.1; K-S, 106.8; 3-C (Third Grade-Constant), 110.5; and 3-S, 108.6. A $2 \times 2$ factorial analysis of variance applied to these data yielded no significant $F$s. The mean IQ for all Ss was 107.8.

*Preliminary training.*—All groups quickly acquired the subgoal seg-

ments, i.e., they learned which panel yielded the marble and which yielded the ball bearing. Eighty-four percent of all Ss attained criterion with no errors and 91% made no more than one error. No child made more than seven errors. The percentages of Ss who attained criterion with no errors were as follows, by groups: K-C, 84%; K-S, 94%; 3-C, 75%; and 3-S, 78%. A $\chi^2$ analysis of the corresponding frequencies revealed no statistically significant differences among the groups.

The major goal segment was also easily learned, i.e., whether the marble or the ball bearing yielded the charm. Except for 1 S who made nine errors and another who made four, all children reached the criterion after no more than three errors. The percentages of Ss who attained criterion with no more than one error were as follows: K-C, 88%; K-S, 88%; 3-C, 69%; 3-S, 84%. A $\chi^2$ analysis of the corresponding frequencies again yielded no significant differences among the groups. It should be noted, however, that in the acquisition of both the subgoal and major goal segments the older Ss made more errors than the younger. This may be another manifestation of the response-shift-tendency reported by Harlow (1959).

*Test trial: initial choice.*—The first component of inferential behavior is the initial choice between A and X, in which A is the inferential choice. The combined results, which appear in Table 1, show a statistically significant age difference in the expected direction ($P < .01$). In fact, on this measure, the kindergarten children perform precisely at chance level.

This part of the results appears to be in conflict with two previous studies (Kendler & Kendler, 1956; Kendler et

TABLE 1

PERCENTAGE OF Ss IN EACH EXPERIMENTAL GROUP WHOSE INITIAL CHOICE WAS A

| Experimental Group | Chronological Age Level | | | |
| --- | --- | --- | --- | --- |
| | 5–6 Yr. | | 8–10 Yr. | |
| | % | $\sigma_P$ | % | $\sigma_P$ |
| Subgoals switched | 50.0 | .088 | 71.9 | .079 |
| Subgoals constant | 50.0 | .088 | 75.0 | .077 |
| Combined | 50.0 | .063 | 73.4 | .056 |

al., 1958) which showed that in groups of children between 34 and 60 mo. of age there were significantly more A than X choices. The studies, however, differed from the present one in two ways. The task used was simpler and the Ss were from higher socioeconomic levels. A more recent study (Kendler & Kendler, 1961) used a procedure and socioeconomic sample comparable to the present ones and obtained similar results. In that study, which had as Ss children between 30 and 65 mo. of age, the number of A and X choices was almost exactly equal. It is therefore suggested that the apparent conflict in results is due to the roles played by the difficulty of the task and the intelligence of the Ss. To some extent this explanation of the discrepancy between the studies is supported by the findings of the present study (see Table 2) that the selection of A is related to *mental* age. It is evident that there was, in general, an increase in the percentage of A choices with increasing MA. This is, however, an ad hoc analysis, not tested for statistical significance; consequently the generality of the results is limited. They do, nevertheless, support the implication of the CA results, namely that inferential behavior is symptomatic of an important developmental process.

*Test trial: integration response.*—If after his initial choice, S inserted a subgoal into the G aperture, he was considered to have made an integration response. If the subgoal thus utilized was B, the integration re-

TABLE 2

PERCENTAGE OF A CHOICES AS A
FUNCTION OF MA

| MA | Chronological Age Levels | | | | | |
|---|---|---|---|---|---|---|
| | 5–6 Yr. | | 8–10 Yr. | | Both | |
| | % A | $\sigma_P$ | % A | $\sigma_P$ | % A | $\sigma_P$ |
| 36–71 | 40.0 | .098 | 100.0 | — | 42.3 | .097 |
| 72–95 | 48.3 | .093 | 50.0 | .145 | 48.8 | .078 |
| 96–119 | 66.7 | .192 | 75.0 | .097 | 73.1 | .087 |
| 120–143 | 100.0 | — | 71.4 | .111 | 77.8 | .098 |
| 144–167 + | — | — | 88.2 | .079 | 88.2 | .079 |

sponse was correct. If it was Y, the integration response was incorrect. Each of these integration responses could be further subdivided into those that occurred with no unnecessary responses intervening between initial and goal responses (direct) and those that occurred after one or more unnecessary responses (indirect). The intervening responses ranged from making only one unnecessary response, e.g., pressing the X button, to repeating almost the entire training sequence, e.g., pressing A but leaving B in trough, then pressing X and leaving Y in trough, then taking both subgoals out and setting them into the same position they occupied during training of B-G, then finally taking up B to drop into the G aperture. (The number of unnecessary responses in indirect solutions may be a fruitful subject for analysis in the future, but for the present the entire gamut was treated as an entity.) Finally, S had the option of making no integration response at all.

Table 3 presents the results for all Ss in the C groups, divided into these five categories. Since it might appear that the superiority of the older children of this measure is merely a reflection of the greater proportion of correct initial responses, the results were also analyzed using as a base only Ss who made an initial A choice, with the following effect. Twelve percent of the kindergartners and 67% of the third-graders who made an initial A choice made direct correct integration responses. This difference is significant ($P < .001$).

There are several conclusions to be drawn from the data in Table 3. One is that, as was the case with the initial choice measure, the integration measure yields marked differences in the performance of the two age levels. Note that the method used by those kindergartners who do attain a correct solution is primarily indirect. Descriptively speaking, such inference was more like trial-and-error than insightful behavior. On the other hand, a very large majority of the older Ss made a correct integration response, and their solutions tended to be more direct than indirect.

The other conclusions deal with the relations between the first and second components of the inferential sequence. Since many Ss, particularly among the kindergartners, who made an A choice either made no integration response at all or interposed other responses before completing the sequence, it may be concluded that the stimulus presented by the "B subgoal

| Chronol. Age (Yr.) | Kind of Integration | | | | No Integration Response |
|---|---|---|---|---|---|
| | A-B-G + (Correct) | | X-Y-G − (Incorrect) | | |
| | Direct | Indirect | Direct | Indirect | |
| 5–6 | 6.2 | 43.8 | 6.2 | 6.2 | 37.5 |
| 8–10 | 50.0 | 37.5 | 3.1 | 3.1 | 6.2 |

in its trough" does not necessarily produce the B-G behavior segment. On the other hand, it is also clear that when an integration response does occur it is much more likely to be a B-G response than a Y-G response. This result could be due to the presence of the B stimulus since it was to this stimulus that the integration response was trained. However, in the Constant Subgoals groups the appearance of the B stimulus depends on an A response. The data have shown that, at least among the older Ss, the A choice is not a matter of chance. It must be due to some problem solving activity, probably some covert system of responses. It is therefore equally possible that the B-G integration may be attributed to the same response mechanism that led to the correct initial choice.

In order to sort out whether the salient influence on the integration response is related to the correctness of the initial choice or to the correctness of the subgoal stimulus, the percentages of direct integration responses under four different conditions are presented in Table 4. Two of these conditions are drawn from the data of the Constant Subgoals groups reported above. The other two conditions are drawn from the Switched

Subgoals groups whose second component results have not yet been reported. In this table R refers to the initial response and S to the subgoal stimulus.

When the proportions of direct integration responses, correct (+) and incorrect (−), under the various conditions are compared for the two age levels, it is apparent that there are too few Ss in the kindergarten groups to yield any trends. The older Ss' behavior, however, varies in an interesting and statistically significant way. A $\chi^2$ analysis of the corresponding frequencies yields a P between .02 and .05. Third graders are most likely to integrate the two components when both their initial choice and the subgoal are correct. When one of these elements is incorrect, the probability of an integration response decreases; however, either element by itself does lead more than one-third of the older Ss to integrate. When neither element is correct, the probability of integration is minimal.

In order to interpret these results it is necessary to examine more closely the nature of these elements. It is to be expected that S+ is more likely to produce integration than S−. But from Table 4 it can be seen that S+ by itself is not sufficient to explain

TABLE 4

PERCENTAGE OF Ss AT EACH AGE LEVEL WHO MADE DIRECT INTEGRATION RESPONSES AS A FUNCTION OF THE CORRECTNESS OF THE INITIAL RESPONSE AND SUBGOAL

| Subgoals | Condition Prior to Integration Response | Type of Integration | Age Level | | | |
|---|---|---|---|---|---|---|
| | | | 5–6 Yr. | | 8–10 Yr. | |
| | | | $N_T$ | % Integr. | $N_T$ | % Integr. |
| Constant | R+S+ | A-B-G+ | 16 | 12 | 24 | 67 |
| Switched | R−S+ | X-B-G+ | 16 | 3 | 9 | 44 |
| Switched | R+S− | A-Y-G− | 16 | 0 | 23 | 35 |
| Constant | R−S− | X-Y-G− | 16 | 12 | 8 | 12 |

integration at either age level. Among the younger $S$s very little integration occurs even under S+ conditions. Among the older $S$s, S+ is important but is not as adequate by itself as when in combination with R+. In fact R+ is so important that it can effect some integration even when it is combined with S−.

To understand why S+ by itself is not the adequate stimulus for integration, it must be realized that it does not stand for the B subgoal in isolation. It actually symbolizes a total stimulus complex of "B in a trough of one of the side panels." During training this particular stimulus compound was associated with picking up B and then returning it to $E$. There are quite a few $S$s, particularly among the kindergartners, who, during the test situation, do just that. It is another stimulus compound, namely B and Y in front of the B-G panel, that is associated with picking up B and dropping it into the aperture. The data indicate that the generalization from the former stimulus compound to the latter is more likely to occur in 8–10 yr. olds than in 5–6 yr. olds. In the older age group it is more likely to occur after a correct initial choice than after an incorrect one.

When these results are represented in terms of S-R associations (Kendler & Kendler, 1962), it becomes possible to describe the psychological difference in the behavior of the two age groups as well as to suggest the relevant mechanisms for the inferential behavior reported in this study. There appear to be three important characteristics of this behavior. One is that making the correct initial choice in an inferential solution depends on the ability of $S$ to generate a covert anticipatory response to the major goal and to respond appropriately to this cues. Such ability, according to this study, is positively related to age (within

the limits tested). A second characteristic is that inferential solution depends on a "short-cutting" process in which some overt S-R sequences previously learned must drop out. The younger $S$s had difficulty in this respect. Repeating some of the responses practiced during training prevented many from exhibiting direct inferential solutions. The third characteristic of inferential behavior is that it is not governed completely by external events. If it were, then it would be expected that the availability of B following a response to X should yield as much integration as when it follows a response to A. Since it does not, one can conclude that the integration response must be influenced by some internal stimulus component associated with the correct initial choice. This source of stimulation is sufficiently potent to compete successfully, in some cases, with external stimulation when there is a conflict between the two. This three-part characterization of inference is not offered as a theory. Its function is to analyze inferential behavior into more fundamental processes so that each may be investigated independently.

## Summary

Children of two age levels, namely 5–6 and 8–10 yr., were presented with a task that required the linkage of two out of three discretely acquired segments of behavior. The solution consisted of making an initial choice between two of the segments (one correct and the other incorrect) and then integrating the product of that choice with the third segment. The solutions could be direct (inferential), i.e., the goal achieved without any unnecessary responses. They could be indirect, i.e., the goal achieved after the repetition of previously acquired but presently irrelevant behavior segments. The findings were: (a) Older $S$s made significantly more correct initial choices than younger $S$s. (b) About half of the younger $S$s ultimately reached the goal, but their method was primarily indirect. Almost all of the older $S$s achieved solution and a majority of them by direct inferential means. (c) In inferential solutions the integration of the subgoal and major segments was a joint function of the relevance of the external stimulation (as

determined by preliminary training) and the correctness of the initial choice. (*d*) When these two contributing factors were experimentally balanced against each other, it was found that the internal stimulation associated with the correct initial response was about as important a determiner of integration as the relevance of the external stimulus.

## REFERENCES

DUNN, L. D. *Peabody Picture Vocabulary Test (PPVT)*. Nashville, Tenn.: American Guidance Service, 1959.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

HULL, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solutions. *Psychol. Rev.*, 1935, **42**, 219–245.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

KENDLER, H. H., & KENDLER, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem-solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KENDLER, H. H., KENDLER, T. S., PLISKOFF, S. S., & D'AMATO, M. F. Inferential behavior in children: I. The influence of reinforcement on incentive motivation. *J. exp. Psychol.*, 1958, **55**, 207–212.

KENDLER, T. S., & KENDLER, H. H. Inferential behavior in children: II. The influence of order of presentation. *J. exp. Psychol.*, 1961, **61**, 442–448.

# THE SEMANTIC MEDIATION OF EVALUATIVE MEANING [1]

FRANCIS J. DI VESTA AND DONALD O. STOVER

*Syracuse University*

Within the theoretical framework of Osgood's (Osgood, Suci, & Tannenbaum, 1957) theory of meaning the *sign* of an object is a primary symbol assumed to evoke a representational mediating response that is some part of the total behavior emitted by the organism when stimulated by the object itself. This response produces distinctive cues, mediating behavior that would otherwise not have occurred in the absence of previous association of the object with the word. The representational elements $(r_m\text{-}s_m)$ correspond to the meaning of the sign, and because of the cue or stimulus components the sign can be conditioned to other stimuli initially lacking in meaning. Upon conditioning such symbolic stimuli are designated *assigns*. On the basis of the congruity principle distinctive responses may be produced by novel stimuli if the assign, within appropriate contextual arrangements, is used to label other novel stimuli (Osgood et al., 1957).

Typical studies of this process are those in which meaning of verbal stimuli have been

changed by simple conditioning (Staats, Staats, & Biggs, 1958) and shown to generalize to synonyms of those stimuli (Staats, Staats, & Heard, 1959). Rhine and Silun (1958) demonstrated that development and strength of a concept-attitude were affected by the amount of reinforcement. Eisman (1955) provided reinforcements for a word and then associated that word with various colored objects, thereby strengthening the probability that the object would be chosen by the S. Using a Treatments × Ss design, Osipow (1960) provided evidence that a color-name associated with either a positive or negative evaluative word resulted in corresponding changes in preference for a nonsense figure when the color-name was associated with the figure. Since significant changes also occurred in the control group his results were inconclusive. Di Vesta (1962) compared the effects of reinforcing a neutral color-name with the effects of attaching the neutral color-name to a positive evaluative sign in the first stage of the mediation process. It was demonstrated that both procedures were effective in changing preferences for a nonsense figure when the color-name was subsequently associated with the figure.

The primary concern in the present study was to compare the effect of two different experimental designs commonly used in the studies summarized above. These experiments were intended to extend the findings of the previous investigations by testing two hypotheses: (*a*) that association of a neutral symbol with several signs, all of which represent similar polarities of the evaluative dimension of meaning, will result in movement in semantic space of the assign corresponding to the connotative meaning of the signs associated with it; and (*b*) that labeling a neutral stimulus object (nonsense figure) with the conditioned assign will result in an evaluation of that stimulus object corresponding to the acquired mean-

ing of the assign, independent of whatever effects may occur from acquired distinctiveness of cues resulting from labeling alone.

Several distinctions between the designs used in the present study and those used in previous investigations may be summarized: A Treatments × Ss design with repeated measurements on each S was employed in Exp. I, and a two-factor design with repeated measurements on each S was employed in Exp. II. Both experiments provide for control of learning-how-to-learn and warm-up effects. The effect of meaning was controlled by using experimental treatments in which neutral evaluative words were conditioned to the assign. In the Treatments × Ss experiment two additional controls were used, one for the use of a label without experimentally acquired meanings, and another for the evaluation of the stimulus object without the use of labels. And, finally, in both experiments the assign was associated with a nonsense figure as the stimulus object rather than with another verbal symbol.

## EXPERIMENT I

### Method

*Design.*—The general procedure used in the experiment was a modification of the mediated generalization paradigm. In Phase 1, the three sets of nonsense syllables were, respectively, conditioned to words with positive, neutral, and negative evaluative meanings. In Phase 2, the Ss learned to name each of three different nonsense figures with one of the three conditioned nonsense syllables, a fourth figure with an unconditioned nonsense syllable, and Ss experienced a fifth figure without naming it. Thus, one of the figures was associated with a positive-conditioned nonsense syllable (positive treatment), a second with a neutral-conditioned nonsense syllable (neutral treatment), a third with a negative-conditioned nonsense syllable (negative treatment), a fourth with an unconditioned nonsense syllable (control-UNS), and a fifth

with no name attached to it (control). In Phase 3, the Ss rated each of the figures on three semantic differential scales.

*Materials.*—Three decks of 5 × 8 in. cards, with 24 cards in each deck, were used in the verbal conditioning phase. One positive, neutral, or negative evaluative word was printed in block letters 1¼ in. high on each card. Examples of the eight positive evaluative words are RIGHT, CLEAN, BRAVE, and SMART; examples of the eight neutral evaluative words are MIDDLE, USUAL, AVERAGE, and MEDIUM; and examples of the eight negative words are WRONG, WICKED, FILTHY, and STUPID. The negative and positive evaluative words were antonyms. Eight different words were used for each of the polar positions. The order in which the cards were arranged in the deck was determined at random with the restriction that each polar position was represented twice in every block of six words and that no more than one meaning occurred twice in succession in each block of six words. The same words were used in the second and third decks and differed from the first deck only in the order in which the words appeared.

Three sets of semantic differential rating scales were used. All scales were measures of evaluative meaning. The first was a practice set to help Ss understand the procedure. The Pretty-Ugly and the Wise-Foolish scales were used. The second set, comprised of the Cruel-Kind, and Wise-Foolish scales, was used to determine whether Ss' meanings for the nonsense syllables were changed in Phase 1. The third set, made up of Good-Bad, Pretty-Ugly, and Like-Dislike scales, was the measure used in Phase 3 to determine the evaluation of the figures. It should be noted that none of the scales duplicated any of the specific words used in the verbal conditioning procedure. All scales were five-point scales, with points 4 cm. apart. The S was permitted to check any place on the scale that he felt best represented his judgment. The intensity of S's response was obtained by measuring the distance, from the extreme end of the scale to his check mark, in centimeters, 1 cm. represented extreme negative ratings, 9 cm. marked the midpoint of the scale, and 17 cm. represented extreme positive ratings.

Five nonsense figures were used, constructed according to the procedure described by Attneave and Arnoult (1956) as Method I for angular shapes with closed contours. Each figure was approximately 3½ × 4 in. and was made of heavy white oaktag paper painted gray. These figures were mounted on a white oaktag square with sides 5½ in. long.

*Procedure.*—The Ss were first introduced to the task and to the method of using the semantic differential scales by having them rate three sets of two pictures (similar to those in the Stanford-Binet Intelligence Scales) on the Pretty-Ugly scale to determine their ability to make these evaluations. In addition, S rated examples of behavior such as "a boy crossing the street without looking," and fictitious characters such as "Donald Duck" on the Wise-Foolish scales. All Ss proceeded succcessfully through this task.

A procedure similar to that described earlier by Di Vesta (1961) was used in Phase 1. The Ss were instructed that they were to learn the meanings of three words (nonsense syllables) that they had not heard before. Nine nonsense syllables were used throughout the experiment, 1 S receiving any three at random from among these, for example, PID, LOM, and SUD might have been used for 1 S and CIT, POH, and LIM for another S. The polar position of evaluative meaning (positive, neutral, or negative) associated with a particular nonsense syllable was randomly arranged within the total experiment in order to balance evaluations that might have been based on associative characteristics of the syllables. (In Exp. II the syllables were found to have no more than chance effects.) The S was then given the first deck of cards and instructed to look at the top card and indicate which of the nonsense syllables it defined. If S thought, for example, that "plain" was a definition for LOM, he was to say "LOM is plain." The correction procedure was used. After S responded, E gave the correct response. If S was correct, he went on to the next card; if incorrect, he repeated the correct association before proceeding to the next card. This procedure was continued until S reached the criterion of responding correctly to two blocks of six words in each. This criterion was used since it signified that S responded correctly to each polarity at least four times with different words. If S did not reach this criterion after proceeding through the three decks twice, he was eliminated from the experiment. At the conclusion of this phase, S rated each of the nonsense syllables (assigns) on the Cruel-Kind and the Wise-Foolish rating scales.

In Phase 2 S was told that the syllables he had just learned about were also the names of certain figures. The S was shown five nonsense figures and told that since they were unlike any other figures that he may have seen, they could not be called by names like rectangles, triangles, and the like. The S was further instructed that four of the figures did have names which he was to learn, and that the names corresponded in three cases to the syllables he had just learned, while the fourth figure also had a name (the unconditioned nonsense syllable) but that he had not learned it, and that the fifth figure had no name. The S was then shown the figures, one at a time, in random order until he could name all labeled ones correctly in four successive presentations without error. A procedure similar to that used in the first stage was used in reinforcing correct responses. In order to randomize any effects of initial preference for the figure, pairing of assigns and figures was varied among Ss.

In the third phase, S was asked to rate the figures, one at a time. The order in which figures were presented to a particular S was determined at random. Each figure was rated on three semantic differential scales, Good-Bad, Pretty-Ugly, and Like-Dislike.

In order to increase the level of motivation S was told that he would earn a toy if he participated in the experiment. At the conclusion of the session he was allowed to select his choice of trinkets from among crayons, baseball picture cards, balloons, and the like.

*Subjects.*—The Ss were 24 children from one fifth-grade class in an elementary school. However, 4 Ss were eliminated from the analysis because they failed to meet the criterion for learning in Phase 1 and 3 other Ss were eliminated because they failed to use the rating scales appropriately. Thus, all analyses were based on an N of 17. The Ss' CAs were 9 and 10 yr.

## Results

An average of 42.70 ($SD = 21.79$) pairings, excluding criterial trials, of nonsense syllables with signs was required to learn the evaluative meanings of the assigns and an average of 28.25 ($SD = 17.63$) pairings was required to achieve the criterion in learning to associate the assign with the nonsense figures. The means and SDs for ratings of the nonsense syllables immediately after conditioning, on the Cruel-Kind and Wise-Foolish scales were compared.[2] Since no more

[2] Two tables in which are presented the means and SDs of assign ratings for each group in Exp. I and II, one table summarizing the learning data for Exp. II, and one table

than two scores overlapped in any one of the comparisons of the distributions of ratings for the three treatments, it was obvious that the differences were significant and no formal tests of significance were made.

Each *S* had been administered all treatments and used three scales in the final ratings of the figures within each treatment. Accordingly, a split split-plots analysis of variance was used.[3] A summary of the Treatments × Scales × *S*s analysis is presented in Table 1. As is evident from the table, the *F*s for the main effects of both Treatments and Scales were significant ($P < .01$). A test of the significance of the differences between the means of the scales indicated that the ratings made on the Pretty-Ugly scale tended to be generally more

---

summarizing the analysis of variance testing the effects of nonsense syllables within each group in Exp. II, have been deposited with the American Documentation Institute. Order Document No. 7261 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[3] The assumption of homogeneity of variance was found to be tenable, via Bartlett's test, for all analyses of variance except where otherwise indicated.

## TABLE 1

ANALYSIS OF VARIANCE OF FINAL RATINGS OF FIGURES: EXP. I

| Source | df | MS | F |
|---|---|---|---|
| Between *S*s | 16 | 24.18 | |
| Within *S*s | | | |
| Treatments (T) | 4 | 195.62 | 5.06* |
| T × *S*s (Error 1) | 64 | 38.70 | |
| Scales (S) | 2 | 63.83 | 7.24* |
| T × S | 8 | 5.92 | <1.00 |
| S × *S*s (Error 2) | 160 | 8.82 | |
| Total | 254 | | |

\* $P < .01$.

negative ($P < .01$) than the ratings made on either of the other two scales. (Similar results were found in Exp. II.) The ratings made on the Good-Bad and the Like-Dislike scales were not significantly different ($P > .05$). The difference in the Pretty-Ugly scale did not interact with treatments as indicated by the *F* ($< 1.00$) for Treatments × Scales interaction. Thus only treatment means obtained by combining the ratings of the three scales to obtain a total score were compared. The overall comparison of treatments using combined scores is represented by the *F* of 5.06 ($P < .01$).

The means and *SD*s of individual scale scores and total scores for Treat-

## TABLE 2

MEANS AND *SD*s OF RATINGS OF FIGURES ON EACH SCALE AND TOTAL SCALES FOR EACH TREATMENT: EXP. I

| Treatment | Good-Bad | | Pretty-Ugly | | Like-Dislike | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Positive | 12.71 | 3.10 | 10.65 | 4.95 | 12.59 | 3.26 | 11.98 | 3.97 |
| Neutral | 9.76 | 4.24 | 7.59 | 3.57 | 9.29 | 4.17 | 8.88 | 4.11 |
| Negative | 7.00 | 5.01 | 6.00 | 4.78 | 7.06 | 5.42 | 6.69 | 5.10 |
| Control-UNS | 8.82 | 2.29 | 7.53 | 3.65 | 7.41 | 3.97 | 8.12 | 3.60 |
| Control | 8.94 | 3.84 | 7.59 | 3.73 | 10.06 | 3.90 | 8.86 | 3.96 |

ments are presented in Table 2. Duncan's multiple range test (Edwards, 1960) was used in testing the multiple comparisons between means. The Treatments × Ss error term was used in making the comparisons. The comparisons of the positive evaluative treatment with each of the four other treatments indicated that all differences were significant ($P < .001$). The comparisons of the negative evaluative treatment with the neutral evaluative treatment and the control without labeling were also significant ($P < .01$). The comparison of the negative evaluative treatment with the control in which only labeling was used was not significant ($.10 > P > .05$). None of the other comparisons among controls and the neutral treatment were significant ($P > .10$).

## EXPERIMENT II

### Method

*Design.*—The primary purpose of this experiment was to replicate a part of Exp. I while excluding carry-over effects from one treatment to the next, a consideration often neglected by previous investigators. Nine groups, in a two-factor design with three repeated measurements in each group, were used. One factor was based on the three polarities (positive, neutral, and negative) of evaluative meaning; the second was based on the use of three different figures. Three figures were used to increase the generality of the results. If no differences were found between figures, the results could then be attributed more definitely to the mediation of meaning. In Phase 1 the nonsense syllable was associated with signs having positive-evaluative meanings for the positive group; with neutral-evaluative meanings for the neutral group; and with negative-evaluative meanings for the negative group. In Phase 2 the conditioned nonsense syllable was associated with Fig. 1 by one-third of each group, Fig. 2 by another third, and Fig. 3 by the last third of each group. In Phase 3 Ss used the three scales, as in the previous experiment, for rating the figure labeled by the assign.
*Procedure.*—The following modifications were made in the procedure used in Exp. I.

In Phase 1, S learned the meaning for only one nonsense syllable. The same list of words was used as in Exp. I and in exactly the same form. However, in Exp. II S indicated whether or not each of the 24 words was the meaning for the nonsense syllable to be conditioned. Thus, for example, if he was in the neutral group he would say, "LOM is plain" or "LOM is not brave." The E then recited the correct association. If S was correct he turned to the next card, if incorrect he repeated the correct combination and then turned to the next card. Eight of the 24 words, depending upon the experimental condition, were used as the definition of the nonsense syllable in each group. In Phase 2 S then identified and named only three figures; one of the figures was labeled by the conditioned nonsense syllable and the other two figures were named with unconditioned nonsense syllables. The three figures were selected on the basis of pretest data with over 150 children in which these figures had been ranked as the most neutral of six figures.

Three different nonsense syllables were used in each group. Different syllables were used between groups. The fact that the three syllables might have comprised a fourth factor in the total design was inadvertently neglected. However, since equal numbers of Ss were assigned each of the syllables within any one condition, the data were analyzed by separate analyses of variance to determine whether the syllables used within each treatment had other than chance effects on the results. In the three analyses thus made, figures and syllables were fixed factors and scales were repeated measures. In every case, the main effect of syllables and the interaction term in which syllables appeared were not significant ($P > .20$). All but 2 of the 12 Fs so calculated were $< 1.00$.[2] Accordingly, the conclusion was that the selected syllables had no effect on the final ratings made.

In Phase 3, the S rated the figure labeled with the conditioned syllable. The ratings were made with the same three scales used in Exp. I. The figures labeled with the unconditioned nonsense syllables were not rated since these data were not relevant to the hypotheses.

*Subjects.*—The Ss were 81 children from the fifth grade classes of a different elementary school from that used in Exp. I. The Ss were 9 and 10 yr. of age, and were randomly assigned to the nine experimental groups. Twelve Ss had been eliminated; 6 Ss failed to learn the task in Phase 1, and 1 S in the positive group, 4 Ss in the neutral group, and

1 in the negative group failed to use the rating scales properly at the end of the experiment.

## Results

The means and SDs of the learning data as measured by errors and trials to reach the criterion for the three main experimental groups in Phases 1 and 2 were compared.[2] The variances of the learning data in Phase 1 were heterogeneous requiring the use of the Kruskal-Wallis analysis of variance to test differences among the groups. Significant ($P < .01$) differences among groups were found for both measures. Multiple comparisons were tested by the Mann-Whitney $U$ test. The difference in errors between the positive and negative group was not significant ($P > .05$) while the comparisons of each of those groups with the neutral group were significant ($P < .01$). All comparisons on the "trials" criterion were significant ($P < .01$). However, the order of difficulty in learning was not correlated with the order of the final preference rating; therefore, the assumption was made that trials to learn or errors made in learning were not related to changes in preferences, and that any differences in final preference could be attributed to the mediation of meaning. No significant differences were found in the comparisons based on either of the learning measures in Phase 2.

The means and SDs of the assign ratings following the verbal conditioning procedure in Phase 1 were summarized.[2] These data compared favorably with those obtained for Exp. I.

The Treatments × Figures × Scales analysis of variance of the ratings made of the figures in Phase 3 is summarized in Table 3. While the overall Treatment effects were significant ($P < .001$) the differences in Scales ($P < .001$) and the interaction

TABLE 3

ANALYSIS OF VARIANCE OF FINAL RATINGS OF FIGURES: EXP. II

| Source | df | MS | F |
|---|---|---|---|
| Between Ss | 80 | | |
| Treatment (T) | 2 | 382.86 | 10.05** |
| Figures (F) | 2 | 6.33 | <1.00 |
| T × F | 4 | 12.53 | <1.00 |
| Error (1) | 72 | 38.10 | |
| Within Ss | 162 | | |
| Scales (S) | 2 | 78.48 | 8.79** |
| T × S | 4 | 32.98 | 3.69* |
| F × S | 4 | 18.30 | 2.05 |
| T × F × S | 8 | 12.66 | 1.42 |
| Error (2) | 144 | 8.93 | |
| Total | 242 | | |

\* $P < .01$.
\*\* $P < .001$.

of Treatments and Scales ($P < .01$) were also significant. The means and SDs of the ratings of the figures on each of the three scales are presented in Table 4. The overall ratings are also summarized to permit comparison with the data from Exp. I. Data for the specific figures have been combined since the $F$ for differences between figures was $< 1.00$. In the multiple comparisons[4] all scales significantly ($P < .01$) differentiated the positive from the negative group; the Good-Bad and the Like-Dislike scales significantly ($P < .05$) differentiated the neutral from the negative group; and the Pretty-Ugly scale significantly ($P < .05$) differentiated the neutral from the positive group. None of the other comparisons was significant. When the data from the three scales were combined the differentiation among the three groups was more clear-cut and the trends clearly corresponded to the conditioning of the syllables as well as to the data for

[4] The modification of Duncan's multiple comparison test described by Collier (1958) was used in testing the differences between treatment means on each scale.

TABLE 4

FINAL RATINGS OF FIGURES ON EACH SCALE AND TOTAL SCALES
BY EXPERIMENTAL GROUPS: EXP. II

| Group | Scales | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Good-Bad | | Pretty-Ugly | | Like-Dislike | | Total | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Positive | 10.67 | 4.01 | 8.78 | 3.75 | 11.00 | 4.64 | 10.15 | 3.26 |
| Neutral | 9.30 | 3.70 | 6.96 | 3.40 | 9.89 | 4.10 | 8.59 | 3.27 |
| Negative | 4.52 | 4.80 | 6.19 | 4.60 | 6.93 | 4.80 | 5.88 | 3.87 |

Exp. I. The difference between the neutral and positive groups were significant ($P < .05$). The differences in the remaining two comparisons were also significant ($P < .05$).

## DISCUSSION

The results of the two experiments demonstrate that (a) assigns acquire evaluative meaning through association with several signs having specific polarity on this dimension of meaning and that (b) these assigns transfer to affect the ratings of figures associated with them. The changes in preferences for the figures corresponded to the meanings, rather than the specific words, with which the nonsense syllable had been associated. Where such associations had not been made, as in the two control groups of Exp. I, the result was a neutral evaluation of the figure. These findings imply the occurrence of the mediation chain consisting of figure, nonsense syllable, associations, and connotative meaning. However, it is evident that the procedures used impose a severe limitation on this interpretation making an alternative explanation based on simple association principles an equally likely assumption. Thus, once the assign had been formed in the first learning phase it became a "synonym" for the associated words. These words, in turn, may have been directly or implicitly associated with the figure during the second learning phase when the figure was associated with the nonsense syllable. A more convincing demonstration of the media-

tion process would necessitate a comparison of the present experiments with those in which the two learning phases are reversed that is, a comparison of the present experiments with one in which the standard mediation paradigm is employed. These studies are currently being planned and are to be conducted in exactly the same settings in which the present studies were made.

Since the main concern, however, was in a comparison of the two designs, their respective merits and limitations warrant attention. Basically, they yielded the same results and one is, in a real sense, a replication of the other. The intra-individual design appears to be more sensitive in the gross demonstration of the effect than is the individual treatments design. Among the assumptions of the intraindividual design are that there are no order effects; and that there are no carry-over effects from one treatment to another. The first is tenable on the premise that treatments were assigned in random order to Ss. The second, however, is a very likely occurrence in the major phases of Exp. I where Ss learned the three meanings concomitantly. The comparisons between the stimuli conceivably facilitate the distinctions in evaluative response. Although the order of the figure was presented at random, once one figure had been rated a comparison with the others was inevitable; and, since the same three scales were employed in each rating it would also be expected that carry-over effects would occur in the use of the scale.

In both experiments, there were significant differences among the results obtained from the different scales. In all but one instance, that is, in the use of the Pretty-Ugly scale in Exp. II, the preferences for all figures based on each of the scales were in the predicted direction. The total score provided more adequate discrimination between groups in both experiments. It appears that the primary factor in the slight differences among scales may be accounted for in terms of lowered reliability when only one scale is used. The fact that other investigators, e.g., Staats, Staats, & Biggs (1958), have used a single scale with reliable results may merely reflect the ability of adults to use the scales with greater accuracy than children.

Within the limitations discussed above, the present study tentatively suggests the applicability of the representational mediation process to the development of attitudes (Dodge, 1955; Osgood et al., 1957). In terms of the operations of measurement with the semantic differential the meaning of a stimulus object is its allocation in multidimensional space, with attitude defined as the projection of this point onto the evaluative dimension of that space (Osgood et al., 1957). The principal concern is with the common elements of the mediating response of the meanings of the signs. When the assign is associated with evaluative signs, the meaning of the assign will depend upon the mediating response evoked by these signs. The assign may be attached to other neutral stimulus objects to influence the responses evoked by these stimuli. Attitudes are thus characterized as implicitly learned processes with potentially bipolar evaluative properties.

## SUMMARY

This study compared two experimental designs using procedures for the study of the mediation process. In both experiments the first phase consisted of assign development by conditioning signs with evaluative meaning to neutral nonsense syllables. Phase 2 involved associating the assign with a neutral nonsense figure. In the third phase Ss rated the nonsense figure on three semantic differential evaluative scales.

There were 17 fifth-grade children in Exp. I. Each S received five treatments. The dependent variable was the rating of five different nonsense figures. Three of the figures were labeled by assigns previously conditioned in Phase 1 to signs having negative, neutral, or positive evaluative meanings, respectively. A fourth figure was labeled by an unconditioned nonsense syllable and the fifth figure was experienced equally often with the others but was not labeled. Ratings, made on each of the scales, of the figures labeled with conditioned assigns corresponded with the evaluative meaning of signs associated with the assigns. No significant differences were found between ratings made of the figure labeled by the neutral assigns, that labeled by the unconditioned nonsense syllable, or the figure not labeled.

In Exp. II there were 81 fifth-grade pupils. In a two-factor design with repeated measures, independent groups were required to associate signs, varying, respectively, on the negative, neutral, and positive polarities of evaluative meaning, to assigns. Within each group the conditioned assigns were used to label one of three different nonsense figures. The remaining two figures were labeled by nonsense syllables. The results of this experiment corroborated those of the first. The overall ratings, using the combined scores of the three scales, differentiated significantly between the three groups. There was, however, a significant interaction between treatments and scales, although in all but one comparison the differences were in the predicted direction. No significant differences were found between figures. Nor were there significant differences between the nonsense syllables used with any one of the treatment groups.

## REFERENCES

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **63**, 452–471.

COLLIER, R. O. Analysis of variance for correlated observations. *Psychometrika*, 1958, **53**, 223–236.

DI VESTA, F. J. Contrast effects in the verbal conditioning of meaning. *J. exp. Psychol.*, 1961, **62**, 535–544.

DI VESTA, F. J. The effects of mediated generalization on the development of children's preferences for figures. *Child Develpm.*, 1962, **33**, 209–222.

DODGE, J. S. A quantitative investigation of the relation between meaning development and context. Unpublished doctoral dissertation, University of Illinois, 1955.

EDWARDS, A. L. *Experimental design in psychological research.* N. Y.: Rinehart, 1960.

EISMAN, B. S. Attitude formation: The development of a color-preference response through mediated generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 321–326.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

OSIPOW, S. H. The effects of verbal mediation on the modification of children's attitudes. *J. educ. Psychol.*, 1960, **51**, 199–207.

RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept-attitude as a function of consistency of reinforcement. *J. exp. Psychol.*, 1958, **55**, 524–529.

STAATS, A. W., STAATS, C. K., & BIGGS, D. A. Meaning of verbal stimuli changed by conditioning. *Amer. J. Psychol.*, 1958, **71**, 429–431.

STAATS, A. W., STAATS, C. K., & HEARD, W. G. Language conditioning of meaning using a semantic generalization paradigm. *J. exp. Psychol.*, 1959, **57**, 187–192.

# REVERSAL AND NONREVERSAL SHIFTS IN CONCEPT FORMATION USING CONSISTENT AND INCONSISTENT RESPONSES

MARTIN HARROW [1] AND ALEXANDER M. BUCHWALD

*Indiana University*

Several previous card-sorting and block-sorting experiments have compared the speed of making reversal and nonreversal shifts (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Mayzner, 1956; Kendler & Kendler, 1959). Typically, a reversal shift has involved learning two successive sorting tasks, with correct responses for Task 2 being based on the same dimension of the stimuli as on Task 1, but with S being required, literally, to reverse his previous sorting responses. For example, in Task 1, red cards must be placed in Sorting Category A and green cards in Category B, and in Task 2 red cards belong in Category B and green cards in Category A. Nonreversal shifts have involved learning one sorting task and then learning a second sorting task based on some dimension of the stimuli which was irrelevant in Task 1. Previous sorting experiments with college students have found that reversal shifts are learned quicker than nonreversal shifts.

In a number of the experiments in this area (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Kendler, 1959) it has been hypothesized that reversal shifts are learned more quickly because Ss respond to the same dimension of the stimuli as was used previously in Task 1 learning, whereas in a nonreversal shift they are required to respond to a new dimension of the

stimuli. According to this analysis, reversal shifts are learned quickly because the same cues that were used during Task 1 learning are again relevant during the learning of Task 2. The S must merely learn to make different responses to the previously used cues. On the basis of this hypothesis the results in the above experiments have been interpreted as supporting a mediational S-R framework. However, reversal groups in the above experiments, besides using the same dimension of the stimuli that was previously relevant, have also been required to make literal reversals, in which the exact opposite sorting response was required.

If the analysis in terms of the advantages of using the same dimension of the stimuli is correct, then this condition alone should be sufficient to produce facilitating effects. According to the above analysis supporting the mediational S-R framework, it would be expected that a shift which allows S to respond to previously relevant cues (for purposes of simplification and to maintain the traditional terminology used previously, this type of shift will also be called a reversal shift) would be learned quickly even if the new responses that are required are not the exact opposite of the previously learned responses. Thus, after learning a concept based on the *number* of stimulus elements, a reversal group switching to a second *number* concept should learn more quickly than a nonreversal group which shifts to a concept based on the

position of the stimuli, regardless of whether or not exactly opposite sorting responses are required. The present experiment attempted to test this analysis.

An interesting factor which has been involved, in an incidental manner, in some of the previous experiments concerning reversal and nonreversal shifts (Kendler & D'Amato, 1955; Kendler & Mayzner, 1956) was also investigated. This variable is the use of consistent and inconsistent responses. These responses occur in experiments which require Ss to sort response cards into compartments labeled with stimulus cards. Consistent responses occur when response cards are placed with stimulus cards which are similar to them in some respect (e.g., when S is required to sort response cards according to their color and must place red response cards into a compartment labeled with a red stimulus card). Inconsistent responses occur when there is little or no similarity between the stimulus card and its appropriate response cards (e.g., when S is required to sort response cards according to their color, but must put red response cards with a nonred stimulus card).

While consistent and inconsistent responses have appeared in some of the experiments comparing reversal and nonreversal shifts the influence of this variable on the ease of making these shifts has not been directly investigated. It is possible that the relative difficulty of reversal shifts as opposed to nonreversal shifts may be affected by whether the responses required are consistent or inconsistent. The present experiment was designed to investigate the influence of this variable. This was done in a card-sorting situation which varied consistent and inconsistent responses in a

systematic fashion while comparing reversal and nonreversal shifts.

## METHOD

*Subjects.*—The Ss were 83 undergraduate students enrolled in introductory psychology courses at Indiana University. Nineteen of these Ss were eliminated for failure to learn the first concept.

*Materials.*—The experiment was carried out in a four-category card-sorting situation in which response cards were sorted into categories labeled with stimulus cards. A deck of 16 response cards, $3 \times 3$ in. in size and similar in principle to the Wisconsin Card Sorting Test (Berg, 1948) was devised. Small green stars were placed on these cards and the cards varied according to the number and position of the stars on them. There were one, two, three, or four stars placed on each response card. The stars were arranged systematically with all possible combinations of number of stars appearing in each of the four corners of the cards. This made a set of 16 response cards, each of which could be sorted according to the number of stars on it or according to the position of the stars on it.

There were 4 stimulus cards and the Ss were required to sort the 16 response cards with them. These stimulus cards were attached to a black card holder. The first stimulus card had one black line in the upper-left hand corner (Card UL1), the second card had two black lines in the lower-left-hand corner (Card LL2), the third had three black lines in the lower-right hand corner (Card LR3), and the fourth had four black lines in the upper-right hand corner (Card UR4). At the top of the card holder, which the stimulus cards were attached to, was a pocket into which could be placed a response card.

Three different "position" concepts were used (where the correct sorting of each response card depends on the position that the stars are placed on the card) and three different "number" concepts were used (where the correct sorting of each response card depends on the number of stars on the card). The three position concepts were a consistent position concept and two inconsistent position concepts (these latter two were a clockwise position concept and a counterclockwise position concept). The consistent position concept required that response cards which had stars located in Corner X be placed with the stimulus card which had its lines located in the *same* corner, Corner X. For example, response cards having stars in the upper-left-

hand corner were placed with Stimulus Card UL1, which also had its lines located in the upper-left-hand corner. Likewise response cards having stars in the lower-left-hand corner belonged with Stimulus Card LL2, response cards having stars in the lower-right-hand corner went with Stimulus Card LR3, and response cards having stars in the upper-right-hand corner belonged with Stimulus Card UR4. One of the inconsistent position concepts (the clockwise concept) required that response cards which had stars in Corner X be sorted with the stimulus card which had its lines located in the corner which is one position clockwise to Corner X. Thus, response cards having their stars located in the lower-left-hand corner had to be sorted with Stimulus Card UL1, which had its lines located in the upper-left-hand corner (one position clockwise). Likewise, response cards having stars in the lower-right-hand corner belonged with Stimulus Card LL2, response cards having stars in the upper-right-hand corner went with Stimulus Card LR3, and cards having stars in the upper-left-hand corner belonged with Stimulus Card UR4. The other inconsistent position concept (the counterclockwise concept) required that response cards which had stars in Corner X be sorted with the stimulus card which had its lines located in the corner which was one position counterclockwise to Corner X.

The three "number" concepts were a consistent number concept, and two inconsistent number concepts—an $(n + 1)$ concept and an $(n - 1)$ concept. The consistent number concept required that response cards with $n$ stars on them be sorted with the stimulus card which also had $n$ lines on it. Thus, response cards which had one star had to be sorted with the stimulus card with one line (Card UL1). Likewise, response cards having two stars belonged with Stimulus Card LL2, response cards having three stars went with Stimulus Card LR3, and response cards having four stars went with Card UR4. One of the inconsistent number concepts—the $(n + 1)$ concept—required that response cards which had $n$ stars be sorted with the stimulus card which had $(n + 1)$ lines. Hence, response cards which had one star had to be sorted with the stimulus card which had two lines (Card LL2). Similarly, response cards having two stars belonged with Stimulus Card LR3, response cards having three stars went with Stimulus Card UR4, and response cards having four stars belonged with Stimulus Card UL1. The other inconsistent number concept—the $(n - 1)$ concept—required that $S$s sort response cards

which had $n$ stars with the stimulus card which had $(n - 1)$ lines on it.

*Design.*—In the present experiment an attempt was made to test the previous analysis that the facilitative effects of reversal shifts are due to the advantages of using the same dimension of the stimuli when shifting concepts. An attempt also was made to examine the influence of consistent and inconsistent responses on reversal and nonreversal shifts. This was achieved in a three-dimensional factorial design, in which $S$s learned two successive card-sorting tasks. During Task 1 all correct responses were inconsistent with half of the $S$s learning an inconsistent position concept and the other half learning an inconsistent number concept.

For Task 2 $S$s were divided into eight subgroups with half of the $S$s learning a concept involving the same dimension of the stimuli as was previously relevant (reversal shifts) and half learning a concept involving a different dimension of the stimuli than was previously relevant (nonreversal shifts). Half of the $S$s learned a concept involving consistent responses and half learned a concept involving inconsistent responses. Similarly, half of the $S$s learned position concepts and half learned number concepts. This permitted a $2 \times 2 \times 2$ factorial design.

*Procedure.*—The $S$s were tested individually. The instructions read to $S$ indicated that as each response card was shown to him he should point to the stimulus card he thought it belonged with, and that by $E$ telling him whether he was right or wrong he would gradually find out where each response card really belonged.

The response cards were presented randomly and individually to each $S$ by being placed in the slot which was at the top of the card holder. The criterion of learning for both Tasks 1 and 2 was 12 successive correct responses. All $S$s who met the criterion on Task 1 within 160 trials were required to learn Task 2.

The change in the pattern of reinforcement for the second concept was made without informing $S$s. The $S$s who had not learned the second concept within 500 trials were arbitrarily assigned a score of 500 for Task 2 learning.

In order to avoid partial reinforcement of the first concept during learning of the second concept (Buss, 1956; Gormezano & Grant, 1958; Harrow & Friedman, 1958) 4 response cards were eliminated from each set of 16 cards, for each group of nonreversal $S$s. This left 12 response cards in each deck. The same

TABLE 1

NUMBER OF TRIALS TO LEARN FIRST CONCEPT (TASK 1)
AND SECOND CONCEPT (TASK 2)

| Condition | | | | | | Trials to Learn | | | | | |
| Task 1 | | Task 2 | | Type of Shift[b] | N | Task 1 | | | Task 2 | | |
| Response[a] | Concept | Response[a] | Concept | | | Mean | Mdn. | Range | Mean | Mdn. | Range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Incon. | Position | Incon. | Position | Rev. | 8 | 79 | 84.5 | 13–133 | 11 | 5.0 | 2–36 |
| Incon. | Position | Incon. | Number | NR | 8 | 55 | 44.5 | 6–136 | 98 | 46.0 | 20–231 |
| Incon. | Position | Consist. | Position | Rev. | 8 | 62 | 60.0 | 10–127 | 9 | 9.0 | 3–18 |
| Incon. | Position | Consist. | Number | NR | 8 | 71 | 65.5 | 40–133 | 18 | 14.0 | 2–50 |
| Incon. | Number | Incon. | Number | Rev. | 8 | 47 | 38.5 | 25–91 | 14 | 14.5 | 2–30 |
| Incon. | Number | Incon. | Position | NR | 8 | 25 | 24.5 | 3–44 | 155 | 107.0 | 8–488 |
| Incon. | Number | Consist. | Number | Rev. | 8 | 40 | 28.0 | 3–110 | 6 | 4.5 | 1–17 |
| Incon. | Number | Consist. | Position | NR | 8 | 53 | 55.0 | 2–100 | 38 | 15.0 | 4–178 |

[a] Incon. = inconsistent; consist. = consistent.
[b] Rev. = reversal; NR = nonreversal.

4 response cards were also eliminated for the corresponding reversal groups.

## RESULTS

The data concerned with the learning of the first concept are reported in Table 1. Both these results and those for the second concept represent the number of trials to learn the task, excluding the 12 criterion trials. To test for differences in speed of learning between $S$s of different groups learning the same first task, two separate analyses of variances were computed. One analysis of variance compared the four groups initially learning an inconsistent position concept and one compared the four groups initially learning the inconsistent number concept. Using 3 and 28 $df$ in each case for the position and number concepts, the overall $F$s were, respectively, 0.48 ($P > .05$), and 1.36 ($P > .05$). Thus, the data indicate that the four groups of $S$s learning each type of task were equated with each other initially.

The results for learning the second concept are also presented in Table 1. In order to determine whether there were any differences in the speed of

learning the concepts among the eight experimental groups, a $2 \times 2 \times 2$ analysis of variance was carried out. Due to the skewness of the data, a logarithmic transformation was used in place of the raw scores to obtain homogeneity of variance (Edwards, 1950). The results, as can be seen in Table 2, show that the reversal groups learned significantly faster than the nonreversal groups ($P < .001$). Likewise, concepts requiring consistent responses were learned significantly faster than those requiring inconsistent responses ($P < .001$). The sig-

TABLE 2

ANALYSIS OF VARIANCE OF LOG TRIALS
TO LEARN SECOND CONCEPT

| Source | $df$ | $F$ |
|---|---|---|
| Reversal-nonreversal (R-NR) | 1 | 41.32** |
| Consistent-inconsistent (C-IC) | 1 | 17.34** |
| Number-position (N-P) | 1 | 0.95 |
| C-IC $\times$ R-NR | 1 | 7.16* |
| N-P $\times$ R-NR | 1 | 0.36 |
| N-P $\times$ C-IC | 1 | 1.27 |
| N-P $\times$ R-NR $\times$ C-IC | 1 | 1.53 |
| Within groups ($MS$) | 56 | (0.187) |

* $P < .01$.
** $P < .001$.

nificant interaction ($P < .01$) between the reversal-nonreversal and consistent-inconsistent groups, appears to be due to the much slower learning of the two nonreversal-inconsistent groups, as can be seen in Table 1.

Further breakdown of the reversal-nonreversal comparison was done by means of individual $t$ tests. The transformed scores were used and the mean square within groups, obtained from the analysis of variance, was used as the basis for the error term. Two-tailed $t$ tests showed that both kinds of reversal groups learned significantly faster than the comparable nonreversal groups. The reversal groups which shifted concepts within the same dimension of the stimuli, without making a literal reversal of their previous responses (nonliteral reversals), learned significantly faster than the comparable nonreversal groups ($t = 2.65$, $df = 56$, $P < .02$). The reversal groups which both shifted concepts within the same dimension of the stimuli and also made responses which were literal reversals of their previous responses, learned significantly faster than the comparable nonreversal groups ($t = 6.43$, $df = 56$, $P < .001$). The two types of reversal groups did not differ significantly from each other ($t = 1.05$, $df = 56$, $P > .05$). There was a significant difference between the two kinds of nonreversal groups ($t = 4.83$, $df = 56$, $P < .001$).

## Discussion

The quicker learning of reversal shifts as opposed to nonreversal shifts was again found in this experiment. It should be remembered that in the present experiment the label, reversal group, was extended to all groups that learned a second concept which required discrimination according to a dimension of the stimuli that was previously relevant during Task 1 learning. It was found, in all cases, that a second sorting task which was based on the same dimension of the stimuli as the first task was learned quicker than comparable concepts which were not based on the same dimension of the stimuli. This occurred even when the reversal task did not require an exact literal reversal of previous sorting responses. Thus, the data support the previous analysis of Buss (1956), Harrow and Friedman (1958), and Kendler and D'Amato (1955), who hypothesized that a reversal shift is learned quickly because it has the advantage of using a dimension of the stimuli which was previously relevant. In a similar manner the results fit in with a mediational S-R approach.

The data also indicate that concepts requiring consistent responses are learned more quickly than concepts requiring inconsistent responses. These results are not surprising. It seems probable that $S$s have frequently made other similar consistent responses before, in their daily lives, and that in the present experiment they quickly generalized to this particular situation.

The significant interaction between reversal-nonreversal and consistent-inconsistent groups suggests that the relative difficulty of reversal as opposed to nonreversal shifts is affected by whether the concepts used require consistent or inconsistent responses. Due to this, when both consistent and inconsistent responses are used in experiments of this type they should be controlled systematically, whether the interest is in the consistent and inconsistent responses (which in themselves have wide general applicability) or whether they are just used incidentally. In the present experiment it seems appropriate to analyze the significant interaction term with respect to the differences ($P < .001$) between the consistent and inconsistent nonreversal groups. Although the literal reversals involved inconsistent responses and the nonliteral reversals did not, the significant interaction appears to reflect the significant differences between the two types of nonreversal groups. Fitting this interpretation, the means of the two kinds of reversal groups did not differ

significantly from each other. It should also be noted, concerning inconsistent responses, that the groups required to make inconsistent responses were more sensitive to the experimental conditions. Thus, in similar card-sorting experiments it may be advisable, when practical, to use groups making inconsistent responses, due to their greater sensitivity.

## SUMMARY

The present experiment tested the notion that reversal shifts are learned more quickly than nonreversal shifts because they involve responding to a dimension of the stimuli which was used previously. This was accomplished in a four-category card-sorting situation in which both number and position concepts were used. Some concepts required consistent responses and other concepts required inconsistent responses. Sixty-four $Ss$ learned two successive card-sorting tasks. During Task 1 all concepts required inconsistent responses. During Task 2 half of the $Ss$ learned reversal tasks and half learned nonreversal tasks. Also, half the concepts used required consistent responses and half required inconsistent responses.

The results indicated that: (a) All types of reversal tasks (tasks requiring the use of a dimension of the stimuli which was previously relevant) were learned in fewer trials than comparable nonreversal tasks (tasks requiring attention to a different dimension of the stimuli). Thus the previously reported hypotheses were supported. (b) Concepts requiring consistent responses were learned in fewer trials than concepts requiring inconsistent responses. (c) The relative difficulty of reversal shifts as opposed to nonreversal shifts is affected by whether consistent or inconsistent responses are used, i.e., there was a significant Reversal-Nonreversal × Consistent-Inconsistent interaction.

## REFERENCES

BERG, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.

BUSS, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

GORMEZANO, I., & GRANT, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

KENDLER, H. H., & D'AMATO, M. J. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, H. H., & MAYZNER, M. S., JR. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

# THE SERIAL POSITION EFFECT OF FREE RECALL [1]

## BENNET B. MURDOCK, JR.

*University of Vermont*

Recently Murdock (1960) has shown that in free recall $R_1$, the total number of words recalled after one presentation, is a linear function of t, total presentation time. Nothing was said about the serial position effect, though this is a well-known phenomenon of free recall (e.g., Deese & Kaufman, 1957). However, given that there is a serial position effect, the simple linear relationship between $R_1$ and t is rather surprising.

In the customary serial position curve of free recall, probability of recall is plotted as a function of serial position. This means, then, that the area under the serial position curve is equal to $R_1$, the number of words recalled after one presentation. If $R_1$ is a linear function of t then it must follow that the area under the serial position curve is also a linear function of t. However, it is not immediately apparent how the serial position curve varies with t in such a way as to maintain this simple linear relationship.

The present experiment was designed as an attempt to determine how the serial position curve varied with list length and presentation rate while still maintaining this linear relationship. Unfortunately, at the end of the experiment it was still not clear how this relationship came about or, for that matter, whether the relationship was even linear after all. The basic reason for this failure was

that the trends which did show up were not consistent enough to justify any clear-cut conclusions. However, a rather definite picture of the serial position curve itself did emerge from the data. Therefore, the present article will be restricted to a quantitive description and attempted explanation of the serial position curve of free recall.

## PROCEDURE

Six groups each had a different combination of list length and presentation rate. These six combinations were 10-2, 20-1, 15-2, 30-1, 20-2, and 40-1; the first number indicates list length and the second number indicates presentation time (in sec.) per item. Thus, 10-2 means a list of 10 words presented at a rate of 2 sec/item. Notice that the first two, middle two, and last two groups were matched for t, total presentation time (20, 30, and 40 sec., respectively).

For each group there were 80 different lists. The lists were constructed by randomly selecting words from the (approximately) 4000 most common English words (Thorndike-Lorge, 1944, G count of 20 and up), except that homonyms, contractions, and archaic words were excluded.

Group testing was used. Lists were read to Ss either at every beat (presentation rate of 1 sec/item) or at every other beat (presentation rate of 2 sec/item) of an electric metronome set at a rate of 60 beats/min.

After each list there was a recall period of 1.5 min. The Ss wrote down as many words as they could remember in any order that they wished. Each recall period was terminated by a verbal "Ready" signal which preceded the start of the next list by 5–10 sec. All groups were given 20 lists per session and four sessions; successive sessions were spaced 2–7 days apart. Nothing was said about rehearsing while the lists were being presented.

In all there were 103 Ss, students of both sexes from the introductory psychology course who were fulfilling a course requirement. Exact Ns by group are shown in Table 1.

RESULTS

The data were first analyzed to determine if practice effects occurred over the four sessions. Analyses of variance showed that there was a significant $(P < .01)$ improvement over the four sessions for Groups 10-2, 15-2, and 20-2; whereas the effect was significant at only the .05 level for Group 30-1 and was not significant $(P > .05)$ for Groups 20-1 and 40-1. However, the largest difference obtained between the best and the worst session for any one group was 1.13 words, and all other intersession differences were less than 1.0 words. Therefore, when this practice effect is divided into four sessions and anywhere from 10 to 40 serial positions its effect on the serial position curves was negligible.

Table 1 shows the means and *SD*s of the number of words recalled per list $(R_1)$. Each mean is based on 80 lists per *S* and from 15 to 19 *S*s per group. As predicted, groups with the same total presentation time did not differ significantly in mean number of words recalled. That is, no signifi-

TABLE 1

MEAN NUMBER OF WORDS RECALLED

| Group | *N* | Mean | *SD* |
|---|---|---|---|
| 10-2 | 18 | 6.39 | 0.76 |
| 20-1 | 16 | 6.87 | 1.16 |
| 15-2 | 19 | 8.25 | 1.40 |
| 30-1 | 19 | 8.82 | 1.98 |
| 20-2 | 15 | 8.53 | 2.08 |
| 40-1 | 16 | 8.24 | 1.08 |

cant differences were found between Groups 10-2 and 20-1 $(t = 1.39)$, between Groups 15-2 and 30-1 $(t = 1.00)$, or Groups 20-2 and 40-1 $(t = 0.48)$.

The serial position curves are shown in Fig. 1. Probability of recall is plotted as a function of serial position. For greater generality, we would also like to use the data from studies by Murdock and Babick (1961) and Deese and Kaufman (1957). In the Murdock-Babick study there were 18 *S*s each tested on 80 different 25-1 lists. In the Deese-Kaufman study there were two groups of 16 *S*s each; one group was tested on 10 different 10-1 lists and the other group was tested on 10 different 32-1 lists. The
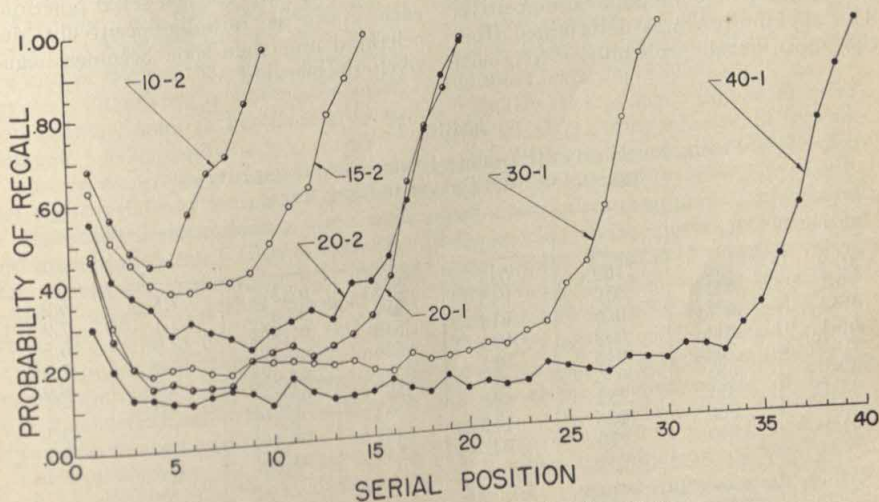


FIG. 1. Serial position curves for the six groups.

serial position data were presented in the original article as Fig. 1 (p. 182) and we read the points from the two curves as accurately as possible. These three serial position curves are shown here as Fig. 2.

We have, then, nine different serial position curves. In general, the curves seem to share certain general characteristics: a marked recency effect, a flat middle section, and a primary effect which is more precipitous though smaller in magnitude than the recency effect. The presence of a flat middle section, or asymptote, is clearest in the 40-1 list (Fig. 1), but becomes less and less obvious as list length decreases. Actually, in the two 10-word lists the primacy and recency curves may have intersected each other before an asymptote has been reached.

More specifically, the recency effect can adequately be described by the Gompertz double-exponential function. As given by Lewis (1960, p. 81) the equation is $y = vg^{h^x}$. Probability of nonrecall $(y)$ was plotted as a function of list length minus serial position $(x)$. Thus, the last word in a list would have an $x$ value of 0, the next to last word an $x$ value of 1, etc. Both $v$ and $g$ were fractional and positive. The asymptote $v$ was determined from the mean recall probabilities averaged
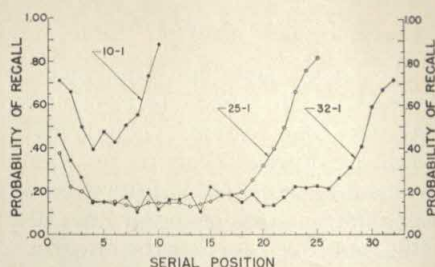


FIG. 2. Serial position curves for 10-1 and 32-1 lists (Deese & Kaufman, 1957) and 25-1 lists (Murdock & Babick, 1961).

over the flat part on each serial position curve. The constants $g$ and $h$ were obtained by a least squares method described by Lewis (1960, pp. 82–88) using the last eight points of each serial position curve (except of the two 10-word lists where only the last four or five points could be used).

The evidence for this conclusion is shown in Table 2 under the $r^2$ column. In all cases the Gompertz equation accounted for more than 95% of the variance, and the mean coefficient of determination $(r^2)$ was 97.79%.

Since in all nine cases $g < 1/e$ the recency effect is consistently an S shaped curve. This characteristic can be seen in the serial position curves of Fig. 1 and 2. Starting from the last serial position, each curve is initially positively decelerated and then soon becomes negatively decelerated.

TABLE 2

VALUES FOR GOMPERTZ DOUBLE-EXPONENTIAL FUNCTION TO FIT
SERIAL POSITION CURVES IN FREE RECALL

| Group | $v$ | $g$ | $y_0$ | $h$ | $r^2$ | $x_i$ | $x_{.95}$ |
|-------|------|------|------|------|--------|--------|--------|
| 10-2 | .548 | .100 | .055 | .574 | 97.3% | 1.5 | 6.9 |
| 20-1 | .852 | .050 | .043 | .596 | 98.8% | 2.1 | 7.9 |
| 15-2 | .622 | .026 | .016 | .518 | 97.8% | 2.0 | 6.5 |
| 30-1 | .814 | .032 | .026 | .546 | 99.3% | 2.0 | 7.0 |
| 20-2 | .730 | .048 | .035 | .552 | 95.9% | 1.9 | 6.9 |
| 40-1 | .885 | .036 | .032 | .557 | 98.7% | 2.0 | 7.1 |
| 25-1[a] | .851 | .134 | .114 | .634 | 98.3% | 1.5 | 8.1 |
| 10-1[b] | .566 | .206 | .117 | .431 | 98.3% | 0.5 | 4.1 |
| 32-1[b] | .840 | .270 | .227 | .644 | 95.7% | 0.6 | 7.4 |

[a] From Murdock and Babick (1961).
[b] From Deese and Kaufman (1957).

The $y_0$ column gives the value of $y$ when $x = 0$. If $y_0$ is subtracted from 1.00 this gives the probability that the word in the last serial position will be correctly recalled. The results for the six groups of the present experiment were very similar to each other, and an analysis of variance of the number correctly recalled showed that the groups did not differ significantly ($F = 1.61$, $df = 5/97$, $P > .05$). The recall probabilities were rather high but they were not 1.00 (and had they been the Gompertz would not be applicable); the corresponding recall probabilities for the Murdock-Babick and Deese-Kaufman data were clearly lower.

The inflection point occurs between the second and third words from the end of the list and appears to be essentially independent of list length and presentation rate. The evidence for this conclusion is given under the $x_i$ column of Table 2, where $x_i = -\ln(-\ln g)/\ln h$ ($\ln$ is log base $e$). That is, $x_i$ is the inflection point, that $x$ value at which the deceleration changes from positive to negative. The $x_i$ values range from 0.5 words to 2.1 words with a mean of 1.57 words. Since the last word in any list has an $x$ value of 0, a mean of 1.57 words places the inflection point midway between the second and third words from the end of the list.

Actually, both Deese-Kaufman curves appear to have inflection points nearer the end of the list than any of the other curves. Otherwise, however, the inflection points cluster rather closely in the range of 1.5–2.1 words.

The recency effect extends over the last eight serial positions and appears to be essentially independent of list length and presentation rate. The evidence for this conclusion is given under the $x_{.95}$ column, where $x_{.95}$ is that value of $x$ at which the curve is 95% down. That is, at this point forgetting is 95% of the asymptotic value. The 95% level serves as a convenient criterion to mark the end of the recency effect.

The mean of the $x_{.95}$ column is 6.88 words or, rounded off to the nearest whole number, 7 words. Except for the Deese-Kaufman 10-1 list all the values seem to be very close to 7 words. Since the $x$ value is 7 words, the recency effect extends over the last eight serial positions.

Another way of indicating the similarity among different lists is by the $h$ column of Table 2. In the Gompertz the constant $h$ determines the rate of change. Since the values of $h$ are all rather similar this indicates that all curves have a similar rate of change, and if they have a similar rate of change all curves should level out at about the same $x$ value if the numerical values of $g$ do not differ too greatly.

The primacy effect appears to extend over the first three or four serial positions. This can be seen in the serial position curves of Fig. 1 and 2, as all of the curves seem to level out at about the third or fourth serial position. The primacy effect is so short-lived that its curve is difficult to describe mathematically. Actually, it may well be exponential. Semilog plots of the first three or four points of the nine curves (using $1.00 - v$ as the asymptote for each curve) gave reasonable approximations to straight lines and the slopes were rather similar to one another. A group curve based on the mean ($y-c$) values of the individual curves was an excellent fit; the rate constant was 0.77 and the intercept was .27 (see Murdock & Cook, 1960). However, the fact that this group curve was based on only three points should make one hesitant about placing too much confidence in it.

Finally, the primacy and recency effects are spanned by a horizontal asymptote. The asymptote is considered to extend from Serial Position 5 up to the last eight serial positions. That is, in a 20-word list the asymptote would extend from Serial Position 5 through Serial Position 12, in a 30-word list from Serial Position 5 through Serial Position 22, etc. That the asymptote is essentially horizontal is suggested by the middle parts of the serial position curves of Fig. 1 and 2.

A close examination of the serial position curves suggests that the trend line may have a small positive

TABLE 3

| Group | $\Delta X$ | Pred. | Obt. | Diff. |
|-------|-----|-------|------|-------|
| 20-1  | 8   | .039  | .100 | .061  |
| 30-1  | 18  | .022  | −.004 | −.026 |
| 20-2  | 8   | .019  | .015 | −.004 |
| 40-1  | 28  | .027  | .038 | .011  |
| 25-1  | 13  | .044  | .036 | −.008 |
| 32-1  | 20  | .032  | .046 | .014  |

slope rather than a zero slope. However, this positive slope could be due to the fact that the recency effect is only 95% down; i.e. 5% of the effect remains to exert an effect on the (allegedly) horizontal asymptote. The proper test of this conclusion, then, is to determine whether the obtained increment (if any) is greater than the increment attributable to the 5% remaining from the recency effect.

The following analysis deals only with lists of 20 words or more; the 10 and 15 word lists could not be used because there were too few points. For each of the six lists the obtained increment was found by fitting a least squares regression line to the asymptote, determining its slope, then multiplying the slope by $\Delta x$ where $\Delta x$ is the difference between Serial Position 5 and the seventh-from-last serial position ($\Delta x = 8$ for the 20-word list, $\Delta x = 13$ for the 25-word list, etc.). The expected increment was found by obtaining the predicted $y$ value from the Gompertz equation for the two values of $x$ (Serial Position 5 and seventh-from-last serial position), then subtracting. For each list the constants shown in Table 2 were used. The predicted and obtained increments are shown in Table 3; the difference between the predicted and obtained increments was not statistically significant ($t = 0.72, df = 5$). Thus, the asymptote does appear to be

horizontal, and the slight positive slope to the curve is no greater than would be expected from the tail end of the recency effect.

## Discussion

We have presented data to show that the serial position curve of free recall is characterized by a rather steep (possibly exponential) primacy effect, an **S** shaped recency effect, and a horizontal asymptote extending between the primacy and recency effect. An idealized curve for a 24-word list is shown in Fig. 3. Its equation is

$$p = 1.00 + .27\ e^{-.77(x-1)}$$
$$- .772\ (.042)^{.555(L-x)}$$

where $L$ is list length and $x$ is Serial Position 1, 2, 3, . . . $L$. The constants for the primacy effect were those of the group curve discussed above while the constants for the asymptote and the recency effect were the mean values of the constants given in Table 2 for the six lists of the present experiment.

The curve of Fig. 3 is an empirical curve, not a rational curve. It is an attempt to describe the serial position effect of free recall quantitatively, not explain it. Not only does this empirical curve represent the nine curves of Fig. 1 and 2 quite well, but also it is consistent with several other sets of data. For one, it agrees with serial position curves for 20-1 lists reported by Deese (1957,
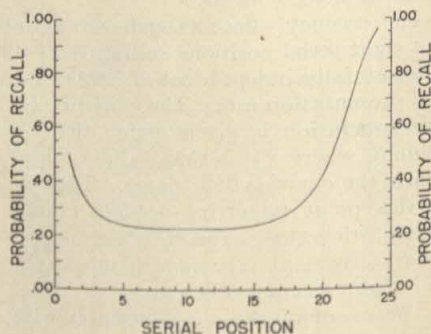
Fig. 3.    Idealized serial position curve for 24-word list.

Fig. 1, p. 580). For another, it agrees well with some unpublished curves culled from several experiments recently reported by Murdock (1960). Finally, the exact same trends are present in some memory-span data reported by Waugh (1960, Fig. 3, p. 75).

However, the empirical curve of Fig. 3 is not in agreement with results reported by Bousfield, Whitmarsh, and Esterson (1958). These authors used 5-, 10-, 20-, and 40-word lists all presented at a rate of 2.5 sec/word, and consistently found the primacy effect more marked than the recency effect. Both Bousfield et al. (1958, pp. 260–261) and Deese (1957, pp. 581–582) suggest that the relatively slow presentation rate may have encouraged rehearsal and thus led to the greater primacy effect. To investigate this possibility, we conducted an additional experiment with 35 Ss using 10 20-2.5 lists. The 20-word length was selected because the curves of Bousfield et al. (1958, Fig. 1, p. 258) seemed to show the most pronounced primacy effect for this length list. As Bousfield et al. (1958) apparently used a somewhat longer recall period we used a 4-min. recall period in this additional experiment; otherwise the procedure was identical with that of the other experiments reported here.

The results of the experiment are shown in Fig. 4. As can be seen, in general the results are quite consistent with the empirical curve of Fig. 1, and in

particular the recency effect is more pronounced than the primacy effect. This experiment clearly shows that the results of Bousfield et al. (1958) are not due to the slower presentation rate per se.

Why did Bousfield et al. (1958) find primacy more pronounced than recency? One possibility is their instructions. Twice in their instructions they told Ss that the words were to be recalled, ". . . in the order in which they occur in your memory." The stress on order may have given Ss a set to recall the words in the order presented, and Deese (1957) has shown that instructions to Ss are an important variable in determining the shape of the curve. A second possibility is the design used. Bousfield et al. (1958) used a counterbalanced design such that each S had only one list at each length. Thus, in effect each list was (to S) of unknown length, and this fact may have encouraged rehearsal in the order of presentation.

In any event, under the conditions of the present experiment there seems little doubt that the serial position effect of free recall is essentially as depicted in Fig. 3. Of course, as Deese (1957, p. 581) has noted, the serial position curve is sensitive to the introduction of experimental variables. However, it has been found that more items are recalled with free recall than with ordered recall (Deese, 1957; Waugh, 1961), so evidently free recall is the preferred, perhaps even the more basic, method of recalling a list of unrelated words.

Finally, why does the serial position curve of free recall take the shape it does? One possible explanation is in terms of short-term proactive and retroactive inhibition. That is, each word in a list is both preceded by anywhere from 0 to $(L - 1)$ other words and followed by anywhere from $(L - 1)$ to 0 other words. Up to a point, the more preceding words the more short-term PI and the more succeeding words the more short-term RI. The PI and RI effects presumably summate to determine the total inhibitory effects.
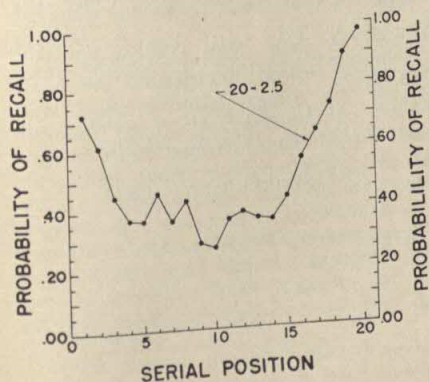
If this explanation is correct, recent



FIG. 4. Serial position curve for 20-2.5 lists.

studies of the short-term retention of individual items should provide an indication of the course of PI and RI to be expected. It has been found that, in short-term memory, PI effects appear to be greatest after about three prior words (Murdock, 1961). This agrees well with the finding that the primacy effect levels out after the first three or four serial positions. In short-term memory RI effects appear to approach an asymptotic value greater than zero (Murdock, 1961; Peterson & Peterson, 1959). This agrees well with the finding of a horizontal asymptote in the serial position curve. Finally, an examination of the RI curve of short-term memory even suggests an S shaped curve (see proportion of correct recalls over different retention intervals, Tables 1 and 3, Murdock, 1961, pp. 619–620). This agrees well with the Gompertz recency effect suggested here. Thus, it would appear that all the main characteristics of the idealized serial position curve shown in Fig. 3 are compatible with the results obtained from the short-term retention of individual items, and these findings lend support to the idea that the serial position curve of free recall is essentially a manifestation of short-term PI and RI effects.

## SUMMARY

This experiment was a study of the serial position effect of free recall. Curves were obtained for 10-2, 20-1, 15-2, 30-1, 20-2, and 40-1 lists, where the first number indicates list length and the second number indicates presentation time per word. On the basis of the available evidence it was concluded that, under the conditions of the present experiment, the serial position curve is characterized by a steep, possible exponential, primacy effect extending over the first three or four words in the list, an S shaped recency effect extending over the last eight words in the list, and a horizontal asymptote spanning the primacy and recency effect. Finally, it was suggested that the shape of the curve may well result from proactive and retroactive inhibition effects occurring within the list itself.

## REFERENCES

BOUSFIELD, W. A., WHITMARSH, G. A., & ESTERSON, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *J. gen. Psychol.*, 1958, **59**, 255–262.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

MURDOCK, B. B., JR., & BABICK, A. J. The effect of repetition on the retention of individual words. *Amer. J. Psychol.*, 1961, **74**, 596–601.

MURDOCK, B. B., JR., & COOK, C. D. On fitting the exponential. *Psychol. Rep.*, 1960, **6**, 63–69.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

WAUGH, N. C. Serial position and the memory-span. *Amer. J. Psychol.*, 1960, **73**, 68–79.

WAUGH, N. C. Free versus serial recall. *J. exp. Psychol.*, 1961, **62**, 496–502.

# THE SCALING OF SUBJECTIVE ROUGHNESS
# AND SMOOTHNESS [1]

## S. S. STEVENS AND JUDITH RICH HARRIS

### Harvard University

These experiments began as an attempt to apply the method of magnitude estimation to a continuum for which the stimulus seemed to have no metric scale, only an ordinal scale of grades of sandpaper, or emery cloth. Unexpected discoveries led on to more engaging inquiries. At the outset, both a ratio scale and a category scale of apparent tactual roughness were determined with 12 grades (grits) of emery cloth. The relation between the ratio scale and the category scale was typical of the relation found on prothetic continua. (For these first results, see Stevens, 1961a; for an earlier related study, see Dudek & Baker, 1956.)

The next study was an exercise in which two students, C. S. Harris and J. P. McMahon, asked 12 Os to judge the smoothness of the stimuli instead of the roughness. The ratio scale of smoothness approximated the inverse, or reciprocal, of the ratio scale found for roughness, and the category scale of smoothness was the reverse, or the complement, of the category scale of roughness. These results resemble Torgerson's (1960) findings when he scaled both the apparent lightness and the apparent darkness of gray papers.

In terms of a linear scale of apparent roughness, it turned out that the stimuli used were bunched rather tightly at the low (smooth) end of the continuum, so much so that the two category scales were almost logarithmic functions of the respective ratio scales. Other studies of category scaling suggest that the nearly logarithmic form found here is an accident of the stimulus spacing, and that if an iterative procedure were used to arrive at a "pure" category scale the curve for roughness would be less curved than

a logarithmic function (Stevens & Galanter, 1957).

Since the available number of emery cloth grits is limited, it is difficult to determine a pure category scale, but when a sample of grits more uniformly spaced in subjective roughness was used, the form of the category scale changed as predicted: it became much less curved than a logarithmic function when plotted against the ratio scale of subjective roughness. Ten Os judged grits 320, 120, 80, 50, 40, 30, and 24 twice each on a seven-point scale. The average judgments were 1.17, 2.17, 3.04, 4.12, 5.17, 6.08, and 6.62. These values determine a line that is straighter than a logarithmic function—a line that is not far from the pure form of the category scale, as evidenced by the tendency of Os to use each category number approximately equally often (Stevens & Galanter, 1957). It appears, therefore, that roughness behaves like a prothetic continuum, and that the pure category scale is not a logarithmic function of the magnitude scale (Eisler, 1962).

From these preliminary studies a surprising fact emerged: magnitude estimations of roughness, and of smoothness, turned out to give fairly straight lines when plotted in log-log coordinates against the grit number of the emery cloths. Another instance, it seems, of the psychophysical power law. (Grit number refers to the number of openings per inch in the screen employed to sift the abrasive particles.) If apparent roughness and its reciprocal, apparent smoothness, are power functions of particle size, it becomes a challenging task to determine more accurately the exponents involved. In the preliminary experiments, each involving 12 Os, the approximate exponents were $-1.5$ for roughness and $+1.2$ for smoothness when measured against grit number. The next problem was to determine

which of these two exponents, if either, is more nearly correct in absolute value.

## APPARATUS AND PROCEDURE

The stimuli were the twelve grits, 320, 240, 220, 180, 120, 100, 80, 60, 50, 40, 30, and 24. It was assumed that these grits, Tri-M-ite brand, met the published standards for abrasives (Horton, 1957) which allow the grain size to vary with a standard deviation of approximately 20% around the nominal size. Two different sets of cloths were used. To a sufficient approximation for the present purpose, the grit number can be regarded as proportional to the reciprocal of the grain diameter.

The stimuli were presented one at a time to $O$, who placed his hand through a cloth-screened opening. He stroked the emery cloth twice with his index and middle fingers. Two samples of each of the 12 grits were presented in a different irregular order to each $O$.

Two experiments used the method of magnitude estimation, and one used cross-modality matching, as follows.

*Experiment 1 (with assigned modulus).*— Grit 100 was presented first and $O$ was told to call it 10. Of the 20 $O$s, 10 first judged roughness and on a later date judged smoothness. The other 10 reversed the order.

Written instructions were given to each $O$. When roughness was being judged, the instructions were:

I am going to present a series of surfaces that vary in roughness. Your task is to tell me how rough they feel by assigning numbers to them. The first will be the standard roughness, which we will call 10. Your task is to assign numbers proportional to your subjective impression. Use whatever numbers seem appropriate—fractions, decimals, or whole numbers. For example, if a surface feels 3 times as rough as the standard, say 30; if it feels ½ as rough, say 2, etc. Try not to worry about being consistent; try to give the appropriate number to each surface regardless of what you might have called some previous surface. In feeling the surfaces, draw your index and middle fingers twice across each surface as it is presented.

When the task was to judge smoothness, the words "smooth" and "smoothness" were substituted for "rough" and "roughness" in the preceding instructions.

*Experiment 2 (no assigned modulus).*—The first stimulus presented was different for each $O$, and, instead of there being a modulus called 10, the instruction was to call the first stimulus "any number you think appropriate." Roughness and smoothness were judged on different days by each of 10 $O$s.

*Experiment 3 (cross-modality matching).*— The $O$ adjusted the intensity of a band of noise (500 to 5000 cps) until its subjective magnitude appeared to match the subjective magnitude of the roughness (or smoothness) of each grade of emery cloth. Two matches were made to each of the 12 cloths in an irregular order by 10 $O$s. One group of 10 $O$s matched loudness to roughness; another group of 10 $O$s matched loudness to smoothness.

The loudness was controlled by a "sone potentiometer" (two 2000-ohm potentiometers, ganged and cascaded). An additional attenuator in series enabled $E$ to keep $O$'s adjustments of the sone potentiometer more or less centered in the usable range. The voltage across the PDR-8 earphones was measured with a vacuum-tube voltmeter.

## RESULTS

The geometric means of the results of the 20 $O$s who judged roughness and smoothness in terms of an assigned modulus (Grit 100 called 10) are shown in Fig. 1. The slopes of the two straight lines are equal but of opposite sign $(+1.4$ and $-1.4)$. This is what is called for by the reciprocal relation between roughness and smoothness. On the other hand, the points do not always fall close to the lines. In general, there is a tendency for both functions to be slightly concave downward.

It is also of interest that the standard called 10 at the beginning of each run was judged less rough when it was presented again as a stimulus to be judged (geometric mean = 6.43). When the task was to judge smoothness, the standard called 10 was later judged more smooth (geometric mean = 13.03). A similar adaptation—if that is what it should be called—was also noted in the preliminary experiments.
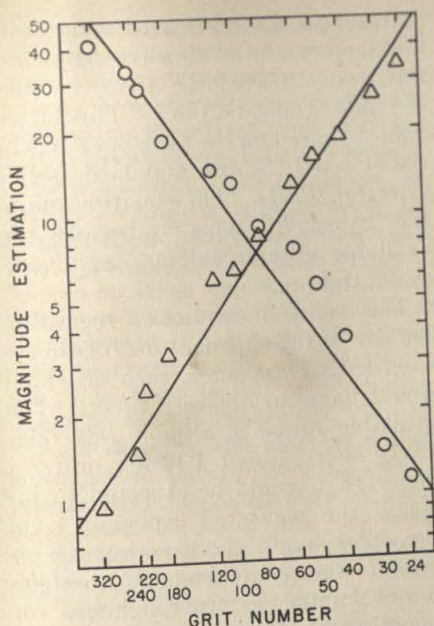
FIG. 1. The geometric means of the estimations of roughness (triangles) and smoothness (circles) are plotted against grit number in log-log coordinates. (Each point is based on 40 judgments—20 Os.)
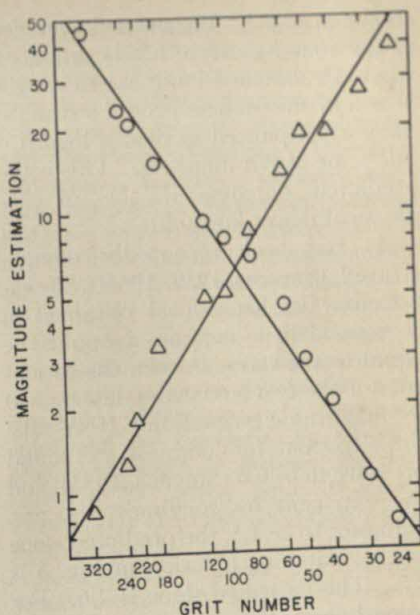


FIG. 2. Similar to Fig. 1, except that in these experiments on roughness (triangles) and smoothness (circles) no standard stimulus was designated by E. (Each point is based on 20 judgments—10 Os.)

Experiment 2, with no assigned modulus, gave the results shown in Fig. 2. The points (geometric means) lie closer to the straight lines and the slopes are ±1.5, a slightly greater absolute value than in Exp. 1. Allowing each O to choose his own modulus appears to have produced a better result. This added freedom has led Os to give superior results on other occasions when no standard was used (Stevens, 1956). Other things being equal, it is better in experiments with magnitude estimation to dispense with an assigned modulus and to average by taking the geometric means of the judgments. No prior processing of the data is necessary with geometric averaging.

The results of matching loudness to roughness and to smoothness are
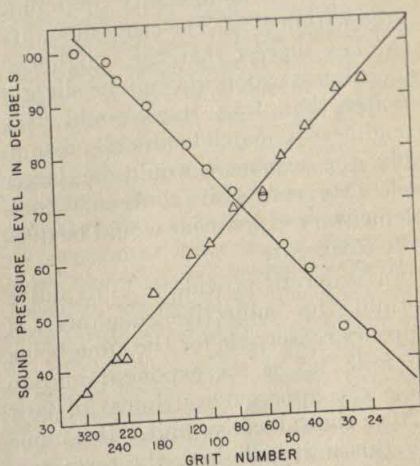


FIG. 3. Sound pressure levels in decibels re 1 microbar produced in the earphones when Os matched the loudness of a noise to the roughness (triangles) and the smoothness (circles) of the 12 grades of emery cloth. (Each point is the decibel average of 20 judgments—10 Os.)

shown in Fig. 3. The decibel averages of the sound pressure levels produced when $O$s matched loudness to roughness and smoothness produce straight lines when plotted against a logarithmic scale of grit numbers. This is the predicted outcome if the psychophysical power law holds.

As has been repeatedly demonstrated (Stevens, 1959, 1961b) the exponents (log-log slopes) obtained in cross-modality matches should be equal to the ratio between the exponents of the two modalities determined by magnitude estimation. If we take the exponent for loudness vs. sound pressure to be 0.6 (Stevens, 1955), and the exponent for roughness vs. grit diameter to be 1.5, the predicted slope of the matching function in Fig. 3 is 2.5. The measured slope is 2.6. For smoothness, the measured slope is $-2.6$, as expected. In view of the sources of error and uncertainty in these experiments, it is reassuring that the measured and the predicted exponents in cross-modality matching agree within 4%. The cross-modality matches suggest that the roughness-smoothness exponents may be slightly greater than 1.5. If we could vary roughness to match loudness, presumably the exponent would be larger still (Stevens, 1959) but this complementary experiment would be difficult to execute.

In order to provide a unit and a formula for subjective roughness, it appears reasonable for the time being to take 1.5 as the exponent and to take the apparent roughness of Grit 320 as the subjective unit. If a name for this unit is desired, the term *ruk*, derived from a root cognate of rough, is suggested. Accordingly, the equation for subjective roughness $R$ in ruks, as a function of grit number $G$ becomes:

$$R = 5724G^{-1.5}$$

In terms of the average diameter in millimeters of the abrasive particles, the exponent would be $+1.5$, and the constant would be 106.5. This latter value is based on the sieve openings in the United States standard series (Horton, 1957). The equation omits the "threshold" constant, because the available stimuli did not permit its evaluation.

The main uncertainties regarding the size of the exponent for roughness stem from two causes, each of which would have an opposite effect. The available range of stimuli was relatively short, about 1.12 log units, a factor that would be expected to increase the measured exponent. On the other hand, the experiments involved the matching of numbers (or of loudnesses) to roughness (or smoothness), never the reverse. Experience has shown that this unbalanced design tends to decrease the measured exponent. Whether the two sources of presumed bias have equal as well as opposite effects cannot be told without further experimental analysis.

## DISCUSSION

*Reciprocality.*—It is clear from Fig. 1, 2, and 3 that values obtained in the judgment of smoothness are approximately the reciprocals of values for apparent roughness. It seems likely that any continuum can be judged with at least fair success in terms of its reciprocal, although only a few have been looked at from this point of view. In addition to the continuum roughness–smoothness, there are data on lightness–darkness (of surfaces) and loudness–softness (of noises). One can easily imagine judging the longness or shortness of lines, the strength or weakness of vibrations, the brightness or dimness of luminances, the heaviness or lightness of lifted weights, and so on. The inverting of a continuum is a kind of semantic matter: $O$ is instructed differently, and he tries to

respond with numbers that are inversely proportional to those he would use for the continuum "right side up." At least he tries to report reciprocals provided he understands the semantic rule: a surface that is twice as rough is half as smooth.

Some continua, like roughness, are frequently referred to in terms of the inverse aspect, but for many continua $O$ would probably find it a little strange to be told to judge the reciprocal. Even with smoothness, a few $O$s commented on the difficulty of the task. It seemed easier and more natural, for example, to match loudness to roughness than to match loudness to smoothness.

Even though people may become accustomed to judging in terms of the inverse aspect, there is a sense in which what may be called the degree of stimulation is basically different from the degree of its absence. When he judges the degree of absence of stimulation, $O$ manages, with fair success, to report the reciprocal of his judgment of the strength of the sensory magnitude. A related type of report was tried out by E. C. Poulton who asked $O$s to estimate fractional loudness by varying only the denominator of their report (see Stevens, 1956). As the tones became fainter the numbers became larger in a manner that produced a nice reciprocal relation to the standard results of magnitude estimation. There are many ways of asking $O$ to report the apparent magnitudes of a series of sensory excitations, but it seems hardly likely that all methods of report that are logically equivalent will give equally good results.

It is also important to note that some sensory attributes that may be thought of as opposites are not at all reciprocally related. Warm and cold are two striking examples (Stevens & Stevens, 1960). The exponent for warm is 1.6, for cold 1.0. These seem to be two different sense modalities, not two names for the same continuum, and the stimulus domains of the two continua do not overlap. Presumably experiments could be run in which $O$ told how warm or how neutral the stimulus felt, and the results would follow the reciprocal relation. An analogous experiment could be run with $O$ judging cold or neutral, again with the expectation of obtaining reciprocal functions, but the slopes (exponents) would differ greatly from the slopes obtained with the warm-neutral pair. One must distinguish, therefore, between word pairs that refer to two different continua and word pairs that refer reciprocally to a single continuum. A given temperature is usually *either* warm or cold, but a given surface is *both* rough and smooth.

*The stimuli.*—The stimuli used in these studies were those samples of emery cloth that happened to be available in the laboratory. They appeared to be in good order, and they seemed to have been a good set of stimuli, in the sense that they gave rise to orderly power functions when used in eight different experiments. It is noteworthy, however, that, as the experiments accumulated, the evidence became increasingly clear that there were minor peculiarities in the series of grits, or perhaps in the manner of their adhesion to the cloths. It is not usual to argue from the results of magnitude estimation to a nonuniformity in a manufacturing process, but the evidence for at least one such "defect" seems clear. The cloth with Grit 120 was consistently judged too rough relative to Grit 100. This fact is especially evident in Fig. 2 and 3 and was also clear in the preliminary experiments.

Under the microscope one can see that Grit 120 involves smaller abrasive grains than Grit 100, as indeed it should, but the two cloths differ in the degree to which the particles appear to be immersed in the adhesive. The finer particles (Grit 120) appear to sit higher on the cloth and to present more of their surface to view. To the touch, the two cloths (100 and 120) feel different in a way that can best be described as qualitative. The skin catches occasionally on the finer particles in a way that is not characteristic of the coarser particles. This catching, due presumably to a shallower immersion of the particles in the bonding adhesive, may account for the relatively higher numerical estima-

tions of the apparent roughness of the finer grit.

Whatever the explanation, it becomes clear that particle size is not the only variable in a bonded abrasive that can affect apparent roughness. Nevertheless, the samples used in these experiments seem to have been only minimally contaminated by other factors, for sensed roughness grows as a relatively clean power function of average particle diameter.

## SUMMARY

Preliminary experiments showed that Os can make consistent judgments of tactual roughness and smoothness. The stimuli were 12 grits of emery cloth. Magnitude estimations of roughness and smoothness produced straight lines when plotted (log-log) against grit number. The exponents of these power functions were determined in two experiments with magnitude estimation and one with cross-modality matching against loudness. All three experiments gave results that were power functions of grit number with exponents in the vicinity of $-1.5$ for roughness and $+1.5$ for smoothness. The cross-modality matches also confirmed the exponents determined by magnitude estimation.

## REFERENCES

DUDEK, F. J., & BAKER, K. E. The constant-sum method applied to scaling subjective dimensions. *Amer. J. Psychol.*, 1956, 69, 616–624.

EISLER, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, 3, 88–96.

HORTON, H. L. (Ed.). *Machinery's handbook.* (15th ed.) New York: Industrial Press, 1957. P. 1459.

STEVENS, J. C., & STEVENS, S. S. Warmth and cold: Dynamics of sensory intensity. *J. exp. Psychol.*, 1960, 60, 183–192.

STEVENS, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, 27, 815–829.

STEVENS, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, 69, 1–25.

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, 57, 201–209.

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, 133, 80–86. (a)

STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 1–33. (b)

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, 54, 377–411.

TORGERSON, W. S. Quantitative judgment scales. In H. Gulliksen and S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960.

# AN EVALUATION OF THE ACTIVATIONIST HYPOTHESIS OF HUMAN VIGILANCE [1]

JACK A. ADAMS AND LAWRENCE R. BOULTER

*University of Illinois*

The arousal or activationist hypothesis of human attentiveness holds that the stimuli impinging on $S$ from external sources, or acting from within him, determine alertness through the reticular activating system of the brain (Broadbent, 1958; Fiske, 1961; Frankmann & Adams, 1962; Lindsley, 1957; Malmo, 1959). Studies of human vigilance, or long-term attentiveness for occasional signals in a monitoring task, commonly show a decrement in signal detection as a function of observation time, and Scott (1957) interprets this phenomenon as adaptation to the nearly unchanging stimulation. Frankmann and Adams (1962) have pointed out that vigilance decrement is not as ubiquitous as is sometimes believed, and that investigators have failed to ask why decrement is more prominently associated with simple tasks and is absent or small in complex visual tasks with multiple stimulus sources. In terms of the activationist hypothesis, there appear to be special sources of stimulation in complex tasks.

Efforts to use explicitly the hypothesis in a careful predictive sense for vigilance findings are hampered by an absence of relationships between the locus, amount, and type of stimulation, and measures of molar behavior (Adams, Stenson, & Humes, 1961; Frankmann & Adams, 1962). How-

ever, with the hypothesis firmly rooted in studies of the reticular formation, there is a number of physiological findings that can be used to facilitate a search for the missing definitions and relationships at the molar level. The two experiments reported here reason from physiological findings and seek to identify sources of stimulation that can deter decrement in complex visual tasks. One derivation was that a special source of stimulation in complex tasks can be the proprioceptive stimulation associated with the head and eye movements of scanning the stimulus array and it is based on proprioception collaterals in the reticular formation (Rossi & Zanchetti, 1957; Samuels, 1959). When the stimulus sources have wide spatial separation, head movements would be the most prominent source of proprioceptive stimulation, but even when multiple stimulus sources are quite close together we can still expect proprioceptive stimulation from eye movements alone (Cooper, Daniel, & Whitteridge, 1955; Whitteridge, 1960). The changing retinal stimulation as the head and eyes scan the visual scene is an added source of stimulation.

Another derivation was that response complexity can produce internal stimulation that should decrease the amount of decrement, and its interaction with the proprioception and visual variables was included to illuminate the effects on vigilance of another source of stimulation in complex tasks. Physiologically, complex responses might be expected to

influence alertness through cortical outputs fed back to the reticular system (Adams et al., 1961; Lindsley, 1957; Rossi & Zanchetti, 1957; Samuels, 1959). Adams and Boulter (1960) and Adams, Stenson, and Humes (1961) have shown that a relatively complex decision response of four choices associated with each signal detection will decrease the amount of vigilance decrement.

## EXPERIMENT I

*Method*

*Apparatus.*—Two units of the multiple source vigilance apparatus were used, with 2 *S*s tested at a time. The *S* sat at a desk and faced a semicircular track fixed at eye level to the back of the desk. On the track, and 30 in. from *S*'s eyes, were mounted four small ($1\frac{1}{2} \times 2$ in. front surface) digital display boxes that presented a two-digit number ($\frac{5}{16}$ in. high) which was the critical signal *S* had to detect. Six different numbers were used. The number was always bright enough for easy reading, but the brightness level was intentionally low so *S* had to orient to a stimulus source to see the signal easily. When the critical signal came on, *S* was instructed to report the event as fast as possible by pressing a detection button located 4 in. from his fingertips. A timer started at the onset of the signal and stopped when the detection button was touched, thus giving a response latency score. This simple mode was called *detection responding*. Under another experimental condition called *memory responding*, which is described in more detail below, *S* had a set of six memory buttons in addition to the detection button. These buttons were each labeled with one of the numbers and were arrayed in a semicircle around the detection button, and 4 in. from it. After *S* pressed the detection button as rapidly as possible, he further selected and pressed a memory button, also as rapidly as possible. The latency from the onset of the signal to pressing the memory button was recorded too. An *S* responded with his preferred hand, but he always kept both forearms in an armrest to standardize the position of the arm and hand. Latency measures such as these have been found to have wide variability when the starting position for the hand is not controlled. A microswitch under each armrest turned on a light at *E*'s panel whenever *S* re-

moved an arm from the armrest and, when this occurred, *E* administered a brief reminder over an intercom system. An occasional reminder in the practice session, which preceded criterion sessions, was usually sufficient to standardize this procedure. All signals and intersignal intervals were programed and automatically read and timed by a digital tape reader that fed the two apparatus units simultaneously. Each *S* was in his own experimental room, which had normal illumination. The *E* and his console were located in another building.

*Description of stimulus series.*—Each session was 2.5 hr. long, and for scoring purposes was divided into five 30-min. trials. There were eight signals on a trial, and each signal lasted for 5 sec. The signals were the numbers 10, 20, 30, 40, 50, and 60, and each number occurred at least once on a trial. Two numbers appeared at each of the four stimulus sources on a trial. Intersignal intervals on each trial were .7, 1.2, 2.0, 3.0, 4.0, 5.0, 6.2, and 7.5 min. Both the assignment of intersignal intervals, and of signals to the four stimulus sources, were separately randomized for each trial. By repeated random sampling, a pool of 10 tapes was generated according to these rules.

*Experimental procedures.*—A major hypothesis was that sustained vigilance in complex visual tasks is a function of the proprioceptive and retinal stimulation arising from head and eye movements, and the manipulation of this class of stimulation was by spatial separation of the four stimulus sources. The assumption was that the wider the separation of sources, the greater the head and eye movements, and the greater the stimulation. Four amounts of spatial separation, specified in terms of the difference between the two outermost sources, were 18°, 36°, 72°, and 144°. The other two sources were spaced between the two peripheral ones to give equal separation between all four. Separations of 18° and 36° placed the four sources in a direct field of visual view, and essentially they could be scanned with eye movements alone. The 72° separation required some added head movements for comfortable scanning, and scanning the 144° separation was not possible without head movements.

A second major hypothesis was that response complexity is a source of internal stimulation for high alertness in complex tasks. Two levels of response complexity were used. One was detection responding, which was the simplest, that only required *S* to press the detection button when a signal was

detected. The other was memory responding, which had much more complexity, and it required $S$ to actively employ his immediate memory throughout the 2.5 hr. The $S$ always had to remember the last set of four numbers that appeared at the stimulus sources. As soon as he detected a new number at a source, his response was to press first the detection button and then press the memory button labeled with the number that had appeared at that same source the last time. This requirement for short-term memory was quite similar to running memory span (Pollack, Johnson, & Knaff, 1959), and it was hypothesized that the continuous memory requirement would introduce greater internal stimulation than detection responding.

Two groups of 15 $S$s each were used, with each $S$ participating for five sessions. A mixed analysis of variance design was used (Lindquist, 1953, p. 292). Response complexity was a between-$S$s variable, with one group having detection responding throughout and another group having memory responding. Within-$S$s variables were five trials and four spatial separations, and were administered to all $S$s of both groups. Following a practice session with a randomly assigned spatial separation of sources, each $S$ was given four criterion sessions where each session was a different one of the four degrees of spatial separation. So that $S$ did not have two identical separation conditions in a row, the three separation conditions that were not used in the practice session were equally divided among the $S$s for the first criterion session. The separation condition for the practice session was assigned to the last criterion session, and the remaining two separations were randomly assigned to the second and third criterion sessions. To control for the possibility of some learning of the characteristics of a particular input tape over five sessions, a tape from the pool of 10 tapes was assigned randomly to an $S$ for each session, with the restriction that no tape be used twice for any $S$. Each session was on a different day.

*Subjects.*—The 30 $S$s were male undergraduate students who were paid for their participation. They were randomly assigned to groups.

## Results

The level of signal detection was high, as might be expected with a signal that persisted for 5 sec. Both

detection and memory responding had a detection level of 98%.

Figure 1 shows the main findings. Each $S$'s score was the mean latency of his responses on a trial, and Fig. 1 presents the plot of group mean latencies as a function of trials. The latencies presented in Fig. 1 are for response to the detection button. Latencies increased as spatial separation increased, and memory responding had longer latencies than detection responding. The differences found between detection and memory responding were found to be primarily associated with pressing the detection button, indicating that all recall and decision delays about which of the six memory buttons to press took place before the main detection button was pressed. The difference between the latency to press the detection button and the latency to press the subsequent memory button is the time to move from the detection button to the memory button, and the means of these difference latencies

TABLE 1

ANALYSIS OF VARIANCE IN EXP. I

| Source | df | MS | F[a] |
|---|---|---|---|
| Between $S$s | | | |
| Detection-Memory (C) | 1 | 95.88 | 31.85* |
| Error$_b$ | 28 | 3.01 | |
| Within $S$s | | | |
| Trials (A) | 4 | 1.23 | 13.67* |
| Spatial separation (B) | 3 | 20.08 | 59.06* |
| A × B | 12 | 0.11 | 1.10 |
| A × C | 4 | 0.15 | 1.67 |
| B × C | 3 | 0.05 | 0.15 |
| A × B × C | 12 | 0.07 | 0.70 |
| Error$_1$(w) | 112 | 0.09 | |
| Error$_2$(w) | 84 | 0.34 | |
| Error$_3$(w) | 336 | 0.10 | |
| Total | 599 | | |

[a] Error$_b$ used to test C effect. Error$_1$(w) used to test A and A × C, Error$_2$(w) used for B and B × C, and Error$_3$(w) used for A × B and A × B × C.
* $P < .01$.

were found to be essentially constant for all conditions. These latter data have been omitted here.

Of primary interest was the influence of spatial separation and response complexity on decremental trends, and Fig. 1 shows decrement to be associated with all conditions. If our hypotheses about the effects of spatial separation and response
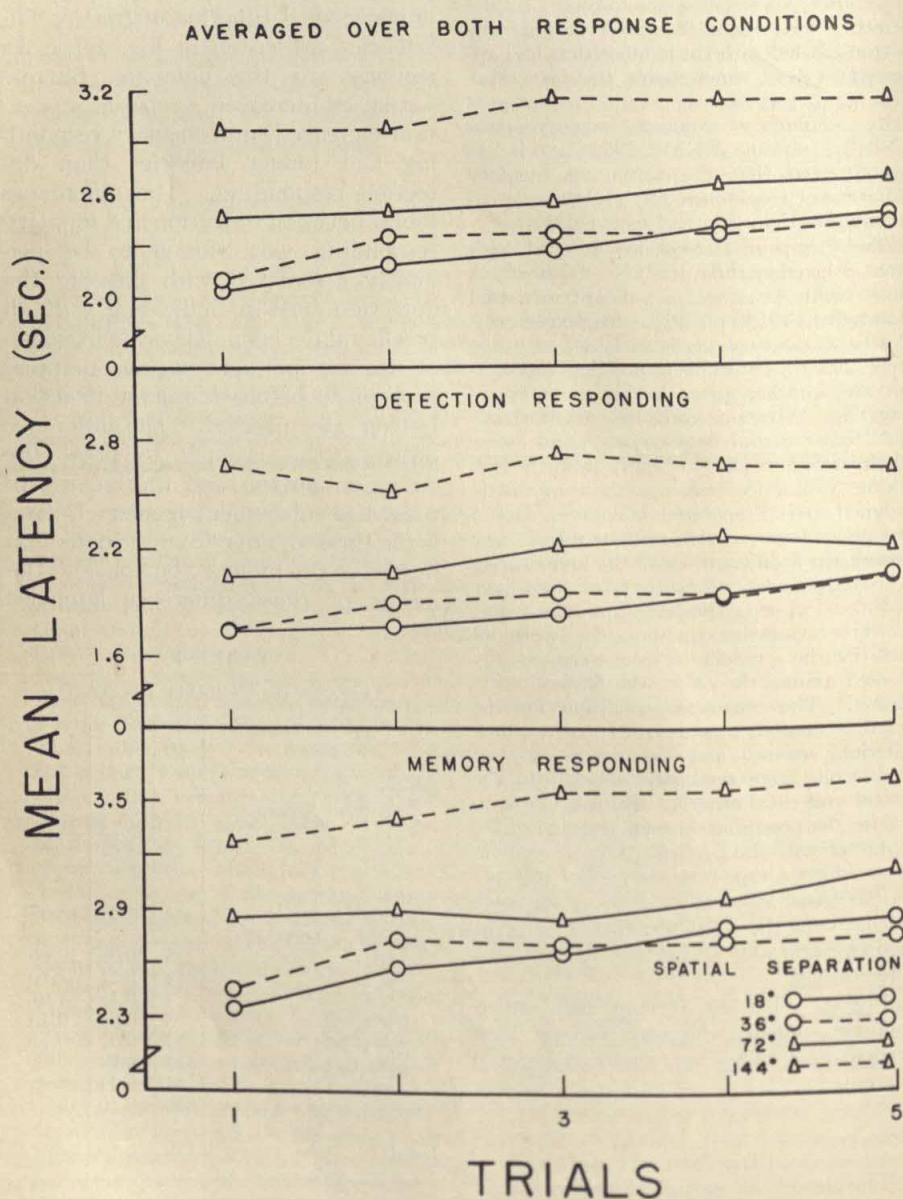


FIG. 1. Mean response latency as a function of the two types of response complexity and the four separations of stimulus sources: Exp. 1.

complexity on decrement are sound, significant interactions should be expected. Less decrement over trials should be found for large spatial separations and for memory responding. The results of the analysis of variance in Table 1 show that the hypothesized interaction effects were not confirmed. All main effects, however, were significant.

## Experiment II

The activationist hypothesis does not clarify the dimensional characteristics of stimulation that influence performance (Adams et al., 1961), and it could be that our operations in Exp. I did not define those particular characteristics of stimulation that deter vigilance decrement. The negative results of Exp. I may have resulted from the manipulation of *amount* rather than the *variety* of stimulation that has been found to be important by other investigators. Sharpless and Jasper (1956) found that stimulus variety induced alertness in the cat, and McGrath (1960) found that visual detection was heightened by varied auditory stimulation such as music, but not auditory white noise. Experiment II was performed to see if *change* in the pattern of head and eye movements could influence vigilance decrement. The plan of the experiment was to establish a pattern of visual observing by procedures analogous to operant reinforcement techniques and then, at a point in a session, require a change in the pattern.

### Method

*Apparatus.*—The same task was used, but it was modified by mounting a small but distinctive neon cue light on top of each stimulus source. In addition to programing critical signals, whose duration was shortened to 2 sec. in this experiment, the tape input mechanism also programed the cue lights

from source to source at 2-sec. intervals. The S was instructed always to watch the cue lights as they changed from source to source because the signal, whenever it occurred, would be coincident with an illuminated cue light. One way of viewing the modified task is in operant reinforcement terms where visually following the changing cue lights is a response class that is rewarded when S sees a number as it first comes on and has the success of an early detection. By faithfully following the cue lights, S tends to optimize his rewards and his performance. In this fashion, we sought to shape a pattern of observing responses that could be subjected to change.

*Experimental procedures.*—The 144° spatial separation was used for all groups to insure an ample observing response. Detection responding was used for all groups. One type of observing response was repetitive (R) where the cue light moved regularly, e.g., Sources 1, 2, 3, 4, 3, 2, 1 . . . etc., so that S scanned regularly back and forth. The other type of observing response was unsystematic (U) where the occurrence of cue lights was random with the restriction that the relative frequency at a source be the same as the R condition. Except for duration being reduced to 2 sec., the occurrences of critical signals on the input tapes were the same as in Exp. I. The shorter signal duration was used to place greater premium on close following of the cue lights because S would soon become aware that he would stand a chance of missing signals if he did not follow the cue lights closely. The length of the session was again 2.5 hr., so the shortening of signal duration slightly increased the length of intersignal intervals.

Four independent groups of 15 Ss each were used. All Ss first had a practice session, and on a subsequent day had a criterion session. Group RU had R scanning throughout except on Trial 3 of the criterion session when it was changed to U scanning. Group R was a control group for Group RU and had R scanning throughout both sessions. If the change of scanning pattern for Group RU influenced vigilance decrement, it would be detectable by comparison with Group R which had no change. As a check on the type of scanning and the effect of its change, another group, UR, had U scanning except when it was changed to R on Trial 3 of the criterion session. Group U was the control for UR, and had U scanning for both sessions.

*Subjects.*—The 60 Ss were university male undergraduates who were paid for their participation. Assignment to groups was random.

*Results*

As in Exp. I, percent detection was high (98%), and the basic score for $S$ again was the mean of his response latencies on a trial. A plot of group means as a function of trials is shown in Fig. 2. Group R, with repetitive visual scanning, had lower mean latencies than Group U with unsystematic scanning, but both groups had decrement. These apparent trends were confirmed in a mixed analysis of variance (Lindquist, 1953, p. 267) where type of scanning (U and R) was the between-$S$s variable, and trials the within-$S$s variable. Type of scanning gave an $F$ ratio ($df = 1/28$) of 19.65 (error $MS = .17$) which was significant at less than the .01 level. Trials had an $F$ ($df = 4/112$) of 9.11 (error $MS = .009$) which also was significant at less than .01. The interaction between these two main effects gave an $F$ ($df = 4/112$) of .44 (error $MS = .009$) which lacked significance at the .05 level.

Trial 3 was the point of variation in scanning for Groups UR and RU, and Fig. 2 shows that the effect of the change was to shift a Trial 3 mean in the direction of its control group.

The $t$ test for independent measures was used on Trial 2 latencies between Groups RU and R, and Group UR and U, to establish comparability of groups before the point of change. The $t$ ratio for the mean difference between RU and R was .05 (error $MS = .004$), and for the difference between UR and U was 1.64 (error $MS = .006$). Both $t$'s lacked significance. Groups RU and U, and Groups UR and R, were then evaluated with the $t$ test at Trial 3, the point of change. The $t$ for the difference between RU and U was .86 (error $MS = .005$), and UR and R was 1.05 (error $MS = .005$). Neither of these two $t$'s was significant at the .05 level. To check if there was a persisting effect in the trial that followed, comparisons of Groups RU and R, and Groups UR and U, were made on Trial 4. The $t$ for the difference between RU and R was .74 (error $MS = .005$), and the $t$ for the comparison between UR and U was 1.50 (error $MS = .006$). Neither $t$ was significant at the .05 level. For all of these $t$ tests, $df = 1/28$.

Interpretation of these statistics must be founded on evidence that the cue lights changed observing behavior. First, cue lights produced lower mean latencies than when they were not used, presumably because $S$ was directed where to look and was often reinforced when he looked there. Detection responding with 144° separation in Exp. I is essentially the same task as given Groups U and R, but without cue lights. Group U had almost twice the speed of responding as the corresponding condition in Exp. I, and Group R was almost three times faster. A second and more direct basis for inferring that cue lights influenced the observing response is shown in Fig. 3, where mean response latency to subsets of stimulus
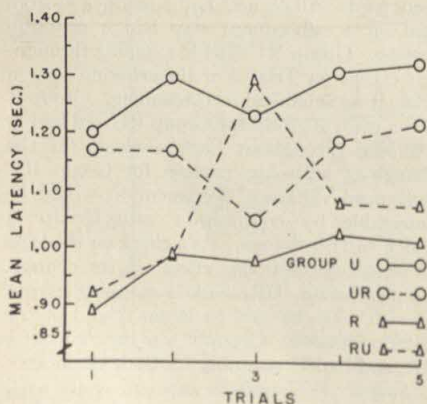


Fig. 2. Mean response latency for the four groups of Exp. II.
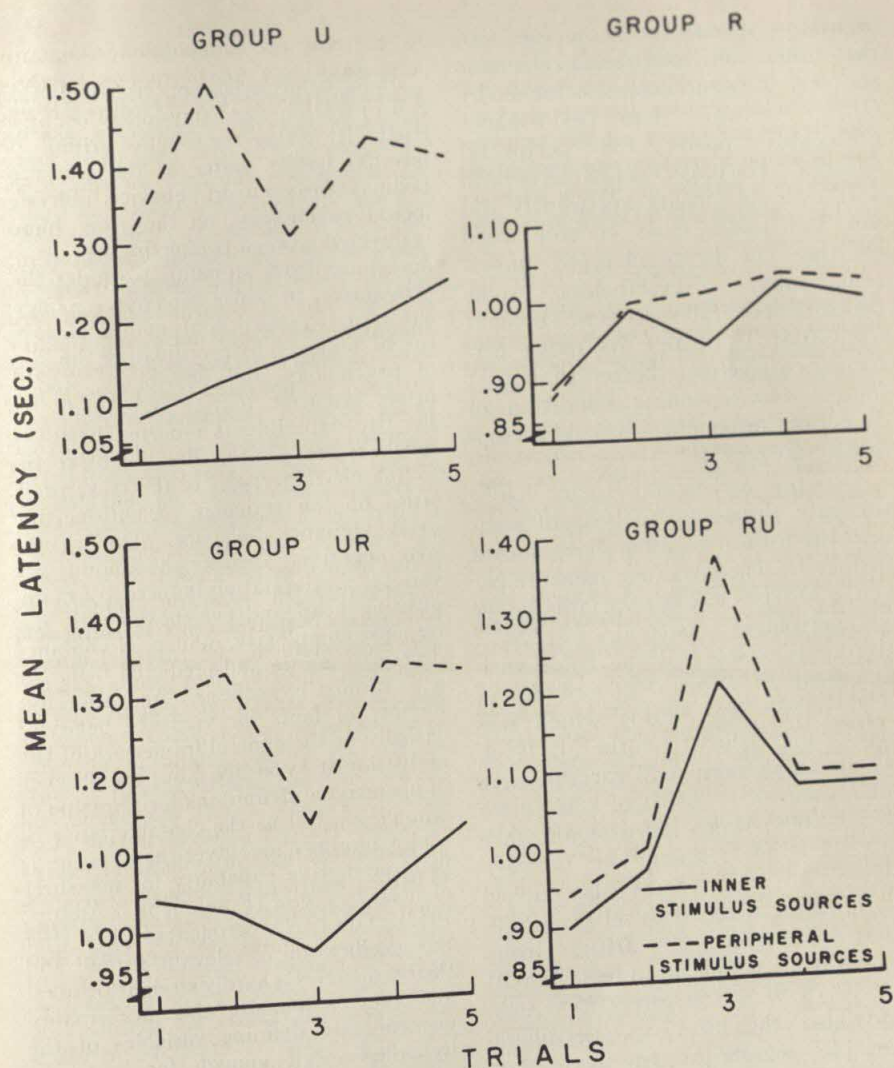
FIG. 3. For the four groups of Exp. II, mean response latency for the inner two and the outer two stimulus sources.

sources is presented. For each $S$ a latency score was computed as the mean response time on a trial to signals of the two inner stimulus sources, and similarly another latency score was computed for signals on a trial for the two peripheral stimulus sources. Figure 3 shows the trial means for each group. If the observing response was under close control and $S$ was systematically following the cue lights, then equal attentiveness to all sources is expected. Figure 3 shows that the expectation was realized for Group R. Not only was their performance level

high, but it was the same for both the inner and peripheral stimulus sources. Less success was achieved by Group U which had performance noticeably poorer for peripheral sources. The larger excursions of head and eye movements apparently did not fall under close control when eliciting cue lights changed unsystematically. Nevertheless, some measure of overall success was evident for Group U because its mean performance level was better than that for the corresponding condition of detection responding and 144° separation in Exp. I.

Group R and Group U each had a separate three-way analysis of variance performed on their data shown in Fig. 3. There was one measure per cell, and trials was one variable, inner vs. peripheral a second, and $S$s a third. For Group R the trials variable had an $F$ ratio ($df = 4/56$) of 6.50 (error $MS = .014$) which was significant at less than the .01 level. The inner vs. peripheral variable gave an $F$ ratio ($df = 1/14$) of 2.44 (error $MS = .009$) which failed significance at .05. The interaction between these two main effects also lacked significance at the .05 level ($F = .70$; $df = 4/56$; error $MS = .010$). Group U had trials significant at less than .01 ($F = 3.76$; $df = 4/56$; error $MS = .025$) and also the inner vs. peripheral variable significant at less than .01 ($F = 38.58$; $df = 1/14$; error $MS = .053$). The interaction between main effects was significant between the .05 and .10 levels ($F = 2.16$; $df = 4/56$; error $MS = .032$).

The plots for Groups UR and RU also are shown in Fig. 3. Change in the patterning of cue lights influenced performance at both inner and peripheral sources, and in directions consistent with those for Groups U and R.

## DISCUSSION

Molar behaviorism concerns itself with S-R operations, and physiology occupies itself with either the effect of stimuli on internal bodily states or the effect of bodily states on overt behavior. Physiological psychology, on the other hand, is directed toward completing the arc of lawfulness from stimulus, to under-the-skin states, to molar behavior. A more substantial network of empirical laws can be achieved by coordinating the laws of psychology and physiology, and a firmer scientific footing is gained. In the two experiments reported here, we sought to move in this direction by trying to link characteristics of the task, properties of the reticular formation, and overt vigilance behavior. Our results were negative. Neither the amount nor variety of stimulation induced by manipulating head and eye movements, nor response complexity, influenced vigilance decrement as we had predicted from the activationist hypothesis. An impressive number of facts for vigilance behavior casually fit the general framework of the activationist hypothesis, but the absence of operational definitions for the type of stimuli, as well as the characteristics of each stimulus class, gives the hypothesis little predictive capability for measures of molar behavior (Frankmann & Adams, 1962). Our two experiments suggest that the specification of relevant stimuli and their properties is hardly straightforward, and they emphasize the problems of operationally defining variables of the hypothesis well enough for it to be thoroughly tested and graduated into a substantive psychophysiological law.

There are several plausible reasons why our tests of the activationist hypothesis failed, and they all illustrate the difficulties of operational definition that must be resolved before the hypothesis can be accepted or rejected. Perhaps our manipulations did not induce differential stimulation of the reticular formation. Adams, Stenson, and Humes (1961) found that number of stimulus sources did not exert a discernible effect on vigilance decrement, and they conjec-

tured that increasing the number of stimulus sources might not have produced an increment in stimulation at the reticular formation. The same reasoning might be applied here. Or, perhaps the operations we used did not sufficiently arouse the reticular formation. We did not have direct physiological measurement of the impact at the reticular formation of these stimuli generated by head and eye movements, but vigorous reticular activation from these operations is certainly an expectation from our current physiological knowledge. Also, the operations of immediate memory might be questioned as a way of inducing cortical-centered stimuli, but the task of keeping the most recent four numbers in mind has a reasonableness about it for a central source of activation. Even if our selection of stimulus classes is granted, we still may not have uncovered manipulations for the relevant dimension of stimuli. Both the memory and the visual scanning tasks in Exp. I may have become repetitive rather early, and decrement could have occurred as a function of adaptation to the steady state of internal stimuli, as Scott (1957) has suggested. Early adaptation may have been a factor in Exp. II where we explicitly tried to introduce variety in our key stimuli. Sustained stimulus variety, whatever its operational definition might be, may prove to be the key to human alertness. Possible directions for inferring relevant dimensions of stimulus variety for visual alertness is the study by McGrath (1960) where meaningful audio stimulation such as music improved detection, and the studies by Adams and Boulter (1960) and Adams, Stenson, and Humes (1961) where four-choice decision responding improved monitoring behavior. McFarland, Holloway, and Hurvich (1942) found that a brief interlude of stretching and conversation with E eliminated decrement in a measure of visual threshold.

The typical vigilance experiment, where marked decrements are found in prolonged monitoring behavior, differs from those presented here in that near threshold stimuli occur at a single stimulus source and percentage of signals detected is the measure. However, there is nothing that would seem to preclude our task with its latency measure as a sound vehicle for testing the hypothesis. Decrement was uniformly obtained but we were unable to control it by manipulating sources of stimulation.

## SUMMARY

The activationist hypothesis contends that environmental and internal sources of stimulation, working through the reticular formation of the brain, are sources of human alertness. Two experiments were performed to identify stimulus determinants of vigilance decrement in a complex visual monitoring task. The $S$'s task was to detect a two-digit number as rapidly as possible when it appeared at any one of four stimulus sources. Response latency was the measure of performance.

Experiment I sought to manipulate response-produced stimulation arising from the stimulation induced by head and eye movements and immediate memory. Amount of head and eye movements was defined in terms of the spatial separation of sources, and the physiological basis of its stimulation value for alertness was hypothesized to be proprioception and visual collaterals in the reticular formation. Immediate memory was defined as a requirement to remember the four numbers that last appeared at the sources and, when a number appeared, to respond with respect to the last number that had appeared at that source. Immediate memory was hypothesized to provide inputs to the reticular formation from cortical areas. No effects of these variables on vigilance decrement were found. Experiment II asked if the negative results of Exp. I were related to amount rather than variety of stimulation being manipulated. Experiment II introduced variety in head and eye movements by training the pattern of visual observing responses and then changing the pattern on a trial. No effects on vigilance decrement were found. Problems of operationally defining the activationist hypothesis were discussed.

## REFERENCES

ADAMS, J. A., & BOULTER, L. R. Monitoring of complex visual displays: I. Effects of response complexity and intersignal interval on vigilant behavior when visual load is moderate. *USAF CCDD tech. Note*, 1960, No. 60-63.

ADAMS, J. A., STENSON, H. H., & HUMES, J. M. Monitoring of complex visual displays: II. Effects of visual load and response complexity on human vigilance. *Hum. Factors*, 1961, **3**, 213–221.

BROADBENT, D. E. *Perception and communication*. New York: Pergamon, 1958.

COOPER, S., DANIEL, P. M., & WHITTERIDGE, D. Muscle spindles and other sensory endings in the extrinsic eye muscles: The physiology and anatomy of these receptors and of their connections with the brainstem. *Brain*, 1955, **78**, 564–583.

FISKE, W. Effects of monotonous and restricted stimulation. In D. W. Fiske and S. R. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey, 1961. Pp. 106–144.

FRANKMANN, J. P., & ADAMS, J. A. Theories of vigilance. *Psychol. Bull.*, 1962, **59**, 257–272.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44–105.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

McFARLAND, R. A., HOLWAY, A. N., & HURVICH, L. M. Studies of visual fatigue. *Publ. Grad. Sch. Bus. Admin. Harvard U.*, 1942.

McGRATH, J. J. The effect of irrelevant environmental stimulation on vigilance performance. Project on Human Factors Problems in Anti-Submarine Warfare, Technical Report No. 6, 1960, Office of Naval Research, Psychological Sciences Division.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

ROSSI, G. F., & ZANCHETTI, F. The brain stem reticular formation. *Arch. Ital. Biol.*, 1957, **95**, 199–435.

SAMUELS, I. Reticular mechanisms and behavior. *Psychol. Bull.*, 1959, **56**, 1–25.

SCOTT, T. H. Literature review of the intellectual effects of perceptual isolation. Report No. HR66, 1957, Department of National Defence, Defence Research Board, Canada.

SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–680.

WHITTERIDGE, D. Central control of eye movements. In J. Field (Ed.), *Handbook of physiology*. Washington: American Physiological Society, 1960. Pp. 1089–1109.

# THE RELATIVE EFFICIENCY OF SEVERAL TRAINING METHODS AS A FUNCTION OF TRANSFER TASK COMPLEXITY [1]

## GEORGE E. BRIGGS AND JAMES C. NAYLOR

*Ohio State University*

In a recent review Naylor (1962) has suggested that the relative efficiency of part- and whole-task training schedules is a function of two variables: *complexity* and *organization.* In his taxonomy, Naylor defines complexity as a function of the *information-processing* demands imposed on $S$ by each component of the task separately, while task organization is specified by the demands placed on $S$ by the interactions or interrelations *between* task components. Thus, task complexity is an intracomponent characteristic, task organization an intercomponent characteristic.

Naylor (1962) has further suggested that the efficiency of a whole-task training method (as compared to a pure-part task method) will increase as task complexity is increased, i.e., there should be an interaction between training method and task complexity in terms of transfer performance. Therefore, those tasks which place greater intradimensional demands upon $S$ should be learned best by whole practice, whereas those tasks having lesser intradimensional task requirements should show a decrease in the relative efficiency of the whole training method.

In tracking tasks, complexity is defined by the order of the machine dynamics through which $S$ attempts to control. Thus, in a multidimensional tracking task, the dynamics of *each* dimension are what collectively determine total task complexity. More specifically, Birmingham and Taylor (1954) pointed out that an increase in dynamics, as from a simple positional (such as setting one's watch) to rate control (such as steering an auto), results in an increase in the information-processing demands placed on $S$, i.e., he must increase the use of derivative information in the tracking error signal to maintain stability in rate task control. In fact, Russell (1951) clearly showed that compared to his transfer function (an analytic expression of his information-processing functions) with a positional control, $S$ will increase his damping (weighting of the first derivative of error in determining a control movement) by a factor of 25 when controlling through rate dynamics. Thus, a more simple transfer function is adequate for position control tasks, but rate dynamics require rate estimates by $S$, thereby increasing the complexity of his task.

Similarly, a further increase in dynamics from rate control to acceleration control results in an even more complex task dimension, since $S$ now must provide a transfer function with the equivalent of two differentiations, one for each of the integral lags present in the control dimension.

The relationship between information-processing demands and control dynamics is discussed in detail by Birmingham and Taylor (1954).

The present study represents an investigation of the task complexity hypothesis. In addition to whole and pure-part task training methods, the present study included groups trained via either a progressive-part or a simplified-whole method.

## METHOD

*Apparatus.*—A three-dimensional compensatory tracking system defined the skill task. The $S$ observed three center-reading meters labeled "heading," "altitude," and "yaw," and he was instructed to maintain null readings on these meters as much of the time as possible. A single input signal was applied to all three dimensions in the form of a simple sinusoid of .03 cps. In the more complex versions of the task (see below) meters showing "altitude rate" and/or "heading rate" were provided. All meters were mounted on a single 10.5 × 19 in. panel which was located approximately 24 in. from $S$, who manipulated a single, three-dimensional control. The control device was a two-dimensional stick where the three control stick movements were left-right (heading), front-back (altitude), and clockwise-counterclockwise rotation of the head (yaw). All three control-display linkages conformed to population stereotypes.

An EASE analog computer provided the task dynamics, as illustrated in Fig. 1. As indicated, there were three levels of control complexity used in this study. Tracking proficiency was measured during each trial in terms of integrated absolute tracking error (average error) for each of the three dimensions separately.

*Design.*—Two independent variables were manipulated in a transfer of training paradigm: transfer task complexity (two levels) and method of training (four methods). The two levels of transfer task complexity were Level II and Level III of Fig. 1. The four training methods were whole-task training, pure-part training, progressive-part training, and simplified-whole task training. There were two groups of $S$s for each of the training methods (one for each of the transfer task complexity levels). The two whole-task groups tracked through Level II and Level III dynamics, respectively, for a total of 10 daily
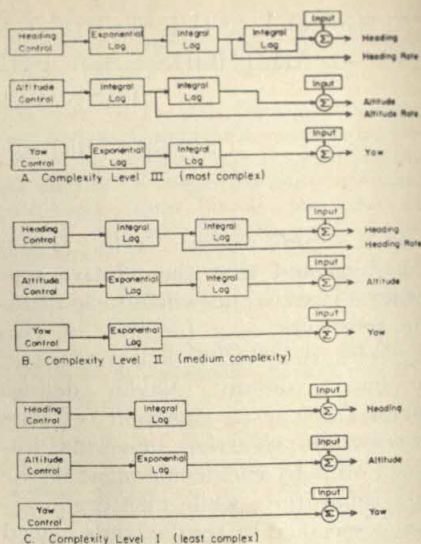


FIG. 1. System dynamics used to obtain various levels of task complexity.

sessions. The remaining six groups received 8 sessions on a training task and 2 sessions on the final transfer task. For three of the six groups the final transfer task involved Level II dynamics, while the other three groups experienced Level III dynamics upon final transfer.

Pure-part training involved four 35-sec. trials per session on each of three separate dimensions. Progressive-part training involved an initial three sessions on a pure-part schedule followed by five sessions during each of which $S$ practiced four 35-sec. trials on each of the three possible pairings of the

## TABLE 1

### EXPERIMENTAL DESIGN OF TRAINING AND TRANSFER CONDITIONS

| Group | Training Condition | Training Task | Transfer Task |
|-------|--------------------|--------------|--------------|
| 1 | Whole | Level III | Level III |
| 2 | Pure-Part | Level III | Level III |
| 3 | Progressive-Part | Level III | Level III |
| 4 | Simplified | Level II | Level III |
| 5 | Whole | Level II | Level II |
| 6 | Pure-Part | Level II | Level II |
| 7 | Progressive-Part | Level II | Level II |
| 8 | Simplified | Level I | Level II |

Note.—See Fig. 1 for a definition of Complexity Levels I, II, and III.

heading, altitude, and yaw dimensions. Of the two groups which experienced the simplified-whole training method, one trained for eight sessions on Level I dynamics (see Fig. 1) and then transferred to Level II, while the other group trained on Level II and transferred to Level III. Table 1 summarizes the training and transfer conditions.

*Subjects and procedure.*—A total of 144 male undergraduates participated in this study. Assignment to the eight groups was based on randomization procedures with the restriction that groups be filled approximately equally throughout the data collection period. At the end of this period there were 18 *Ss* per group. The *Ss* were reimbursed $10.00 for their time.

There were 12 35-sec. tracking trials during each of the 10 sessions. Practice occurred in four-trial blocks with 25 sec. rest between trials within a block and 1 min. rest between blocks. Scoring occurred over the final 30 sec. of each trial to avoid the initial transients in tracking performance; thus, a block of four trials served as the unit of data analysis and represents a behavioral sample of 2 min. for each *S*. During the 12 trials of the initial session, all *Ss* controlled the task *without* the .03-cps input signal; that input was present during all remaining trials. The purpose of this procedure was to facilitate the initial acquisition of skill. All three levels of task dynamics provide inherently unstable systems, and thus the system tends to go out of control quite rapidly for the skill novice; removing the input markedly improved *S*'s chances of "keeping control" over the system during the initial training trials as it provided a less difficult task.

All *Ss* were instructed to keep the meter displays representing heading, altitude, and yaw at their null position at all times, if possible. It was emphasized that control over all three task dimensions was required, and *S* was reminded of this if his performance indicated that he was paying particular attention to one or two displays to the detriment of the other dimension(s). The *S* was informed also of the usefulness of the rate display (if it were present, see Fig. 1).

The performance metric, average error, is actually the average deviation of the error amplitude distribution generated by *S* during each tracking trial. Therefore, the metric describes *S*'s intratrial variability in tracking accuracy. The score was recorded originally in voltage units; however, following data collection all scores were transformed to units of inches of arc, thereby providing a scale matching that of *S*'s tracking display.

## RESULTS

The results from both the training and the transfer sessions are summarized in Fig. 2 (Groups 1–4) and Fig. 3 (Groups 5–8). The tracking error scores were combined for all three dimensions and averaged for each group. In order to provide comparable data points, it was necessary to derive daily averages for the training sessions, i.e., the pure-part and the progressive-part groups experienced all three tracking dimensions during each of the eight training sessions; however, during transfer it was possible to derive comparable data points for four-trial blocks. The use of combined error scores is justified here, since the performance functions on the three dimensions are quite similar in relative rank order and no apparent artifacts are introduced by dealing with "total task" performance in either the training or the transfer sessions.

*Training.*—Several points are of general interest from the training data of Fig. 2 and 3. First, as expected, tracking performance is best during Sessions 1 through 8 for that whole-task group, Group 8, which experienced the least complex system (Level I), and the least proficient performance was attained by Group 1 which experienced the most complex
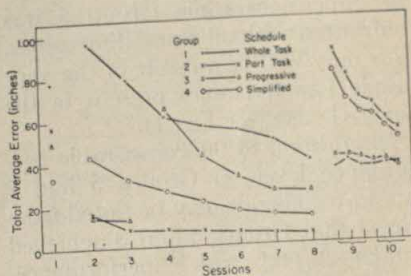


FIG. 2. Training and transfer performance levels for Groups 1–4. (Transfer to Task Complexity Level III.)
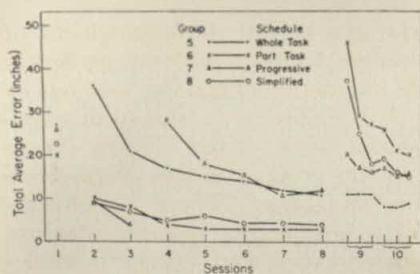
FIG. 3. Training and transfer performance levels for Groups 5–8. (Transfer to Task Complexity Level II.)

whole task (Level III). Both Groups 4 and 5 trained on the same whole task (Level II), and their performance levels (a) are not significantly different one from another (P > .05) and (b) fall intermediate to those of Groups 1 and 8.

It may be recalled that no input signal was employed during Session 1 for any group; thus, these data are set apart in Fig. 2 and 3. Also, it may be recalled that Groups 3 and 7 transferred from a pure-part schedule to practice on pairs of task dimensions on Session 4; therefore, the figures show a discontinuity between Sessions 3 and 4 for Groups 3 and 7. It is interesting to note on Session 4 and thereafter that Group 3 surpasses the whole-task group, Group 1, whereas Group 7 never does exceed the performance of Group 5, even though Group 7 experienced only two tracking dimensions while Group 5 was confronted with all three dimensions. This may well be a result of the very simple yaw dynamics present in the Level II task (see Fig. 1).

In addition to the comparable performance levels of Groups 5 and 7 late in training, it may be noted from Fig. 3 that Groups 6 and 8 achieved similar "total task" performance. Again, this is probably a result of the very simple task dynamics encountered by Group 8 during training. The more complex dynamics experienced by Groups 1–4 would account, therefore, for the clear separation of performance levels for these groups during the latter training sessions (see Fig. 2).

An analysis of variance was applied to the data of Groups 1–3 and Groups 5–7 on Session 8. Groups 4 and 8 were not included, since their task dynamics differed from those of the companion groups. The results of the analysis indicated that training methods ($F = 78.6$; $df = 2/102$), task complexity ($F = 85.2$; $df = 1/102$), and the Methods × Complexity interaction ($F = 28.0$; $df = 2/102$) were statistically significant at $P < .05$. The interaction, illustrated in Fig. 2 and 3, shows that the greatest performance difference between the Level II and the Level III tasks occurred with the whole-task groups, while the progressive-part groups showed less influence of the task complexity variable, and no difference ($P > .05$) was found between Groups 3 and 7. The only other group difference which did not attain statistical significance at $P < .05$ on Session 8 was that between the whole and progressive-part groups on the Level II task dynamics (Groups 5 and 7 as mentioned earlier).

Finally, it may be noted from Fig. 2 and 3 that all groups had approached asymptotic performance levels at or before transfer. This was especially the case for the two part-task practice groups (Groups 2 and 6). If anything, Groups 2 and 6 were overtrained in eight sessions.

*Transfer.*—It would appear from Fig. 3 that during the initial block of transfer trials Groups 6 and 8 performed at a level inferior to Group 5 on Session 2. It should be recalled that the data point for Group 5 on Session 2 is the average for three blocks of trials, and when performance for Group 5 on the first *block* of Session 2 is compared to that of

Groups 6 and 8 on the first block of Session 9, the result is not quite as unexpected. The transfer index $(C_i - E_t/C_i - C_t) \times 100$ was used for this comparison, where $C_i$ is the performance of Group 5 on the first block of Session 2, $C_t$ is the performance of the same group on the first block of Session 9, and $E_t$ is the performance of Group 6, 7, or 8 on the latter block of trials. This index of transfer is particularly appropriate when the raw data are error scores, as in the present case, and the index describes relative improvement in performance as a function of training.

The same index was applied to the data of Groups 1–4 and the results for all groups were: part training = 2% transfer to Complexity Level II, 31% to Complexity Level III; progressive-part training = 75% transfer to Level II, 94% to Level III; and simplified training = 28% transfer to Level II, 44% to Level III. One may note that (a) Group 3 attained a very high level of performance during transfer (94%), while Groups 2, 4, 6, and 8 attained rather poor relative transfer performance levels, and (b) transfer performance was relatively higher on the more complex task than on the less complex task. It is important to note that these transfer data are expressed in relative terms, as indicated above. If examined in terms of *absolute* transfer performance levels, the pattern of results as listed is partially reversed.

This latter point may be illustrated by reference to an analysis of variance which was performed on the data of all eight groups over the entire six blocks of transfer trials. The results are summarized in Table 2 where it is apparent that, as was the case with the training data, there are significant differences among the training methods and between the two levels of task complexity. Further, and of direct

relevance to the above statement on absolute transfer performance levels, the interaction of training methods by task complexity is statistically significant at $P < .05$. Figure 4 contains the group averages which define this interaction. There it is obvious that considerably greater absolute differences occurred between the whole-task group (Group 1) on the more complex transfer task and the groups trained by either pure-part or simplified-whole methods (Groups 2 and 4) than between the comparable groups on the less complex transfer task (Group 5 vs. Groups 6 and 8). However, again it should be emphasized that *relative* comparisons between control (whole training) groups and the experimental groups show better transfer performance on the more complex task than on the less complex task for the latter groups.

The other statistically significant interactions indicated in Table 2 are of interest. The Transfer Blocks × Training Methods interaction indicates a difference in the slopes of the transfer performance curves for the several groups. By reference to Fig. 2 and 3 it may be seen that the slopes for the whole task and progressive-part task groups are very slight, whereas those for the part and

TABLE 2

ANALYSIS OF VARIANCE AS APPLIED TO THE TRANSFER TRIAL BLOCKS OF SESSIONS 9 AND 10 FOR ALL GROUPS

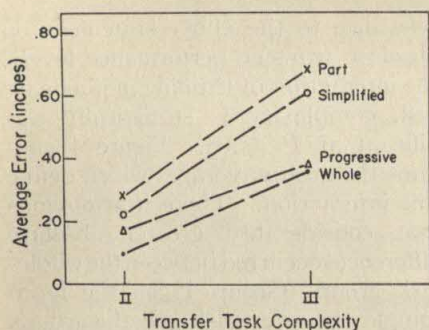| Source | df | MS | F |
|---|---|---|---|
| Training methods (M) | 3 | .764 | 16.6* |
| Task complexity (C) | 1 | 5.410 | 117.6* |
| M × C | 3 | .132 | 2.9* |
| Ss within methods (Ss/M) | 136 | .046 | |
| Transfer blocks (B) | 5 | .122 | 40.7* |
| B × M | 15 | .030 | 10.0* |
| B × C | 5 | .007 | 2.3* |
| B × C × M | 15 | .003 | 1.0 |
| B × Ss/M | 680 | .003 | |

* $P < .05$.

FIG. 4. Transfer task performance averaged over Sessions 9 and 10 for all groups.

simplified-whole groups are rather steep. Thus, while initial transfer performance for the latter two groups was relatively poor, their progress following initial transfer was quite rapid.

The Transfer Blocks × Task Complexity interaction is more difficult to see in Fig. 2 and 3. However, it is a result primarily of the especially high rates of improvement by Groups 2 and 4 on the more complex task. It follows, therefore, that not only was relative transfer higher for these groups on the more complex task, but also their rate of approach to the performance level of the control

TABLE 3

GROUP COMPARISONS OF TRANSFER
PERFORMANCE ON BLOCK 1
(UPPER PART) AND ON
BLOCK 6 (LOWER
PART) FOR ALL
GROUPS

| Cond. | Level III | | | | Level II | | | |
|---|---|---|---|---|---|---|---|---|
| | W | P | P-P | S | W | P | P-P | S |
| W | | .05 | ns | .05 | | .05 | ns | .05 |
| P | ns | | .05 | ns | ns | | .05 | ns |
| P-P | ns | ns | | .05 | ns | ns | | .05 |
| S | ns | ns | ns | | ns | ns | ns | |

Note.—Cell entries indicate statistical significance at $P < .05$ and at $P > .05$ (ns). W, P, P-P, and S identify whole, part, progressive-part, and simplified-whole methods.

group (Group 1) was greater than that for the comparable groups on the less complex system.

Following the analysis of variance, summarized in Table 2, a Duncan multiple range test (Duncan, 1955) was applied to the data for all groups on the first block of transfer trials and on the last block. A summary of these comparisons is provided in Table 3. The upper half of each section of Table 3 represents comparisons made with the initial block of transfer trials, while the lower half is from the last block of transfer trials. These results show that whereas there were differences among groups early in transfer, the relatively short transfer experience with the whole task eliminated all group differences; therefore, the effects of training method were quite transitory.

DISCUSSION

Several findings of the study are of particular interest. First, the relative transfer indices are higher for all three experimental groups on the more complex task. Using this measure of method efficiency, one finds an indication that the part-training methods are actually more efficient when the task is highly complex —a denial of the Naylor hypothesis. Second, using absolute transfer performance (see Fig. 4), one finds that only the progressive-part schedule showed an improvement relative to the whole method. These data lead to the conclusion that both the pure-part and the simplified-whole methods become less efficient training techniques for highly complex tasks and that the whole and some schedule of progressive-part methods emerge as more satisfactory training techniques as task complexity is increased. This substantiates the hypothesis that task complexity and training method interact, with whole training becoming relatively more efficient as complexity is increased. Since the validity of the hypothesis therefore de-

pends upon the measure of transfer that is used, it is not possible to accept or reject without ambiguity.

One possible source for this ambiguity lies in the second task variable mentioned by Naylor (1962). It is likely that an increase in component complexity results in a corresponding increment in the organizational demands of the task. With the three-dimensional task used here, the primary organizational requirements imposed on $S$ consisted of time-sharing demands, i.e., he had to develop an efficient pattern of attention alternation among the three displays and avoid undue emphasis on a single system dimension (component). Thus, as task complexity is increased, $S$ must coordinate his responses to the several task dimensions more exactly to avoid large errors in one or another of the dimensions.

A comparison illustrating the importance of time sharing is that between the simplified-whole, the progressive-part, and the pure-part training methods. If increases in complexity of the system components influence task organization, then the pure-part method should show less effect of task complexity, while the progressive-part and the simplified-whole methods should be affected to a greater extent. This prediction was substantiated by the data: there is no statistically significant difference between Groups 2 and 6 (pure-part) during the training sessions; however, both Groups 3 and 7 (progressive-part) and Groups 4 and 8 (simplified-whole) differ markedly (see Fig. 2 and 3).

Perhaps one of the more intriguing findings related to time-sharing concerns the relationship between the whole and the progressive-part groups during transfer. It would appear that time-sharing can be learned quite well by the progressive-part method, as there were no significant differences between the whole and progressive-part groups at either level on Transfer Block 1 (Table 3). Since, at the most, only practice on pairs of task dimensions was given in the progressive-part method, the acquisition of time-sharing skill may not require practice on the total task.

It would appear, then, that the time-sharing demands of multidimensional control tasks represent a potent variable determining the relative efficiency of training methods. However, this variable accounts for only part of the transfer results. If time-sharing requirements were the only variable of importance, then those groups trained via a simplified-whole method should have been superior to the pure-part groups during transfer. It may be noted from Table 3 that these groups did not differ significantly at $P < .05$; thus, the apparent superiority of Groups 4 and 8 over Groups 2 and 6, respectively, in Fig. 2 and 3 is only apparent—not real. This was a completely unexpected finding of the present research. It was expected on the basis of an earlier analysis by Briggs and Waters (1958) that a simplified-whole training method would provide superior transfer performance compared to that following pure-part training.

Also, Briggs, Fitts, and Bahrick (1958) had previously demonstrated that greater amounts of training resulted in greater amounts of transfer for a simplified training procedure. Since the amount of training provided in this study was greater than that used in the Briggs et al. study, large amounts of transfer for the simplified groups were anticipated.

Therefore, it may be that the benefit accrued by the simplified-whole groups during training (the acquisition of time-sharing skills) was offset in comparison with the pure-part groups by their training on the specific dynamics of the transfer task. Similarity of training and transfer tasks emerges, therefore, as a second major determinant of transfer performance.

It is concluded that some form of progressive-part training will be superior to methods such as pure-part and simplified-whole for the acquisition of skill in a complex, multidimensional task, since the progressive-part method utilizes a training task of high similarity to the transfer task and it also provides an opportunity to develop efficient time-sharing behavior.

Finally, attention is directed to the fact that practice on the whole task during transfer rather quickly eliminated, at least statistically, the differential effects of training methods (see Table 3). Since whole-task training resulted in numerically superior performance throughout training, the only argument for the pure-part or simplified-whole methods rests upon the potential savings in training time on the whole task itself following training. This argument is tenable only to the extent that (*a*) a minimum of whole-task training can equate the effects of the methods in terms of final performance, and (*b*) during transfer following pure-part or simplified-whole training the initial proficiency of *S* is able to meet some criterion of operating safety, i.e., it is conceivable that even after extended training *S* could "lose control" of a complex system and thereby violate safety margins for himself and the system.

The first point above was met in the present study by the very rapid progress of the pure-part and simplified-whole groups during transfer; however, their original transfer performance was very poor, and the pure-part groups, in particular, probably would have been judged "dangerous" had their training and transfer tasks been in a real-life vehicular system.

## SUMMARY

The relative efficiency of four training methods (pure-part, progressive-part, simplified-whole, and whole task) in the acquisition of skill was investigated in a three-dimensional tracking task. Two levels of transfer task complexity were used. There were eight daily training sessions followed by two transfer sessions on the whole task.

The results showed that training via the whole and the progressive-part methods resulted in statistically equivalent performance during transfer for both levels of transfer task complexity. Both the whole and the progressive-part methods resulted in transfer performance which was statistically superior to that of *S*s trained via either the pure-part or the simplified-whole method. Further, as task complexity was increased, the whole and the progressive-part training methods increased the absolute (but not the relative) superiority of transfer performance compared to that with pure-part and simplified-whole methods.

Two factors emerged from a logical analysis of the results to explain the superiority of the whole and the progressive-part methods: first, the transfer (whole) tasks required rather efficient time-sharing skills and these could not be acquired under the pure-part method since, by definition, *S* experiences only one dimension of the task at a time under this training method, and second, training and transfer task similarity was not as high for the simplified-whole groups as it was for those *S*s trained under the progressive-part method; thus, even though the former *S*s could acquire the necessary time-sharing skills, they were "penalized" by training on a task which bore less similarity to the transfer task than was the case for the latter *S*s.

## REFERENCES

BIRMINGHAM, H. P., & TAYLOR, F. V. A human engineering approach to the design of man-operated continuous control systems. *USN Res. Lab. Rep.*, 1954, No. 4333.

BRIGGS, G. E., FITTS, P. M., & BAHRICK, H. P. Transfer effects from a single to a double integral tracking system. *J. exp. Psychol.*, 1958, 55, 135–142.

BRIGGS, G. E., & WATERS, L. K. Training and transfer as a function of component interaction. *J. exp. Psychol.*, 1958, 56, 492–500.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, 11, 1–42.

NAYLOR, J. C. Parameters affecting the relative efficiency of part and whole practice methods: A review of the literature. *USN Train. Dev. Cent. tech. Rep.*, 1962, No. 950-1.

RUSSELL, L. Characteristics of the human as a linear servo-element. Unpublished master's thesis, Massachusetts Institute of Technology, 1951.

# RECRUITMENT, LATENCY, MAGNITUDE, AND AMPLITUDE OF THE GSR AS A FUNCTION OF INTERSTIMULUS INTERVAL [1]

WILLIAM F. PROKASY, JAMES T. FAWCETT [2] AND JOHN F. HALL

*Pennsylvania State University*

White and Schlosberg (1952) and Moeller (1954) report that, within the ranges of 0 and 5 sec., the optimum interstimulus interval (ISI) in GSR conditioning is approximately .5 sec. when magnitude of response is employed as the dependent variable. Bierbaum (1955), in contrast, reports a greater response magnitude near 3 sec.

While it is not clear why his ISI function should diverge from that obtained by White and Schlosberg (1952) and Moeller (1954), Bierbaum (1955) reports, in addition, that recruitment (time elapsing between the onset of a GSR and its maximum) increases with increases in the ISI. Unpublished data from our laboratory suggest, further, that both GSR latency and recruitment modify across extinction trials.

It is the purpose of the present study (*a*) to examine further GSR conditioning as a function of ISI and (*b*) to relate the independent variable (ISI) to three GSR attributes other than magnitude: latency, recruitment, and amplitude. The distinction between magnitude and amplitude first was made by Humphreys (1943) and more recently was discussed by Hil-

gard (1951, p. 528). These authors employed the word *magnitude* to refer to a mean based upon all trials, including those which resulted in no measurable response, while they adopted the word *amplitude* to refer to means derived only from those trials on which a response occurred.

## METHOD

*Subjects.*—The *S*s were 129 men and women enrolled in introductory psychology who volunteered with the knowledge that shock would be employed. Of these, 23 were lost due to equipment malfunction, adaptation to shock, or *E*'s error.

*Apparatus.*—The GSR was measured by means of the Fels dermohmeter. The recording electrodes were .75-in. zinc electrodes set at the base of Plexiglas cups of .125-in. depth. A zinc oxide electrode paste filled the cups and made contact with *S*. The cups were placed approximately 1 in. apart (center to center) on *S*'s left palm with a constant current of $70\mu a$. transmitted between them. Resistance changes were recorded on an Esterline-Angus milliammeter operating at a paper speed of 24 in. per min.

The CS, a 76-db. (re .0002 dynes/cm²) 1000-cps tone, was presented by means of a Grason-Stadler twin oscillator through Permoflux PDR-8 earphones, and was superimposed on a constant background of white noise generated through the earphones at 52 db. The UCS was a shock administered by an ac variac through .5-in. copper electrodes taped to *S*'s right index finger. CS and UCS duration were controlled through Hunter-Brown interval timers.

*Procedure.*—The *S*s were assigned, randomly, to one of five treatment conditions: 0-, .5-, 1-, 3-, or 5-sec. ISI. CS durations for the five conditions, respectively, were: .2, .5, 1, 3, and 5 sec. UCS duration was .2 sec. and began with the termination of the CS in all except Group 0, in which the CS and UCS overlapped.

The $S$ was seated in a sound-shielded room (6 × 10 × 6.5 ft.) which was illuminated by a 75-w. incandescent bulb. After affixing GSR and shock electrodes, $E$ indicated that shock would be administered in increasing amounts, beginning with a barely perceptible level, until it reached a level which $S$ judged to be "highly annoying, but not painful." The mean voltage achieved was 58.7 ($SD$ = 14.5). The $S$ was then instructed that he would receive either the already experienced shock to the finger, a tone through the earphones, or both, but that there was no necessary relationship between the two events. The earphones were positioned and $E$ returned to the outer chamber to initiate training trials. All $S$s received 20 CS-UCS pairings followed by 10 test (extinction) trials. The intertrial interval was varied unsystematically between 25 and 35 sec. Final $N$s in each group were: Group 0, 20; Group .5, 22; Group 1, 22; Group 3, 20; and Group 5, 22.

*Response measures.*—The measure employed on each test trial was the log of conductance change, as measured from the base at CS onset to the point of maximum pen deflection following any response initiation that occurred between 1 and 5.5 sec. after CS onset. To log conductance change was added a constant of 9, thus making all scores positive with a range of from .549 to 3. (Only 3 $S$s yielded scores below 1, and then on only several trials each.) In the event that the pen deflection failed to meet the criterion of a response, a score of zero was recorded. The minimum deflection recorded (as measured by a specially constructed template) was 25, 50, 100, 250, or 500 ohms, depending upon the dermohmeter sensitivity setting required to adjust appropriately to the extent of $S$'s UCR during shock adjustment.

Both magnitude and amplitude were recorded as the mean of log conductance change plus 9 on Test Trial 1 and on each of three blocks of three of the subsequent nine test trials. Magnitude was obtained by including the score of zero on trials in which no response occurred. The amplitude index did not include scores of zero. For example, if $S$'s scores on Test Trials 2, 3, and 4 were 2.5, 0, and 1.1, mean magnitude for those trials would be 1.2 while mean amplitude would be 1.8.

Latency was defined as the time elapsing between CS onset and the occurrence of the first response in the criterion range, while recruitment was defined as the amount of time elapsing between response initiation and response peak. The units of our template were in .03-in. increments which, in conjunc-
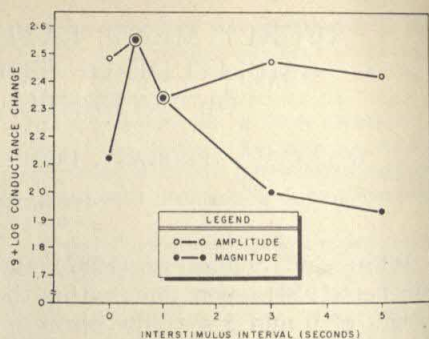


Fig. 1. Amplitude and magnitude as a function of ISI, measured on Test Trial 1.

tion with the paper speed of 24 in. per min., permitted time measurement to the nearest 75 msec.

## RESULTS

*First extinction trial.*—Responses on the first extinction trial were employed as the most direct index of the influence of the ISI variable, as subsequent responses necessarily reflected the added operation of UCS omission. Mean magnitude and amplitude as a function of ISI are shown in Fig. 1. Amplitude means, based only upon those $S$s who responded, are obtained from $N$s of 17, 22, 22, 16, and 17 for Groups 0, .5, 1, 3, and 5, respectively.

Magnitude is greatest at .5 sec. and declines with longer ISI intervals, the decrease over ISI values of .5 to 5 sec. being significant at the .05 level

TABLE 1

LATENCY AND RECRUITMENT MEANS AND $SD$s ON TEST TRIAL 1 AS A FUNCTION OF ISI

| ISI (Sec.) | $N$ | Latency (Sec.) | | Recruitment (Sec.) | |
|---|---|---|---|---|---|
| | | Mean | $SD$ | Mean | $SD$ |
| 0 | 17 | 2.48 | .79 | 2.50 | .72 |
| .5 | 22 | 2.55 | .48 | 3.14 | 1.04 |
| 1 | 22 | 2.58 | .90 | 2.93 | .90 |
| 3 | 16 | 3.11 | 1.22 | 3.23 | .96 |
| 5 | 17 | 2.76 | .98 | 2.87 | 1.82 |

TABLE 2

MEAN LATENCY, RECRUITMENT, AMPLITUDE, AND MAGNITUDE OVER THREE BLOCKS OF THREE TEST TRIALS AS A FUNCTION OF ISI

| ISI (Sec.) | N | Amplitude | | | Magnitude | | | Latency (Sec.) | | | Recruitment (Sec.) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Trial Block | | | Trial Block | | | Trial Block | | | Trial Block | | |
| | | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 0 | 9 | 2.457 | 2.384 | 2.581 | 2.285 | 1.854 | 2.459 | 2.53 | 2.58 | 2.52 | 2.05 | 2.06 | 2.06 |
| .5 | 19 | 2.506 | 2.375 | 2.343 | 2.459 | 2.197 | 2.136 | 2.48 | 2.54 | 2.49 | 2.15 | 2.09 | 2.12 |
| 1 | 18 | 2.451 | 2.400 | 2.391 | 2.415 | 2.223 | 2.203 | 2.47 | 2.58 | 2.75 | 2.61 | 2.39 | 2.36 |
| 3 | 17 | 2.409 | 2.305 | 2.404 | 2.201 | 2.085 | 1.926 | 3.17 | 3.31 | 3.32 | 3.23 | 2.84 | 3.02 |
| 5 | 18 | 2.402 | 2.367 | 2.290 | 2.374 | 2.230 | 2.096 | 2.89 | 2.95 | 3.41 | 2.93 | 3.00 | 2.73 |

Note.—Within-cell $SD$s varied from .18 to .50 for amplitude, from .24 to .80 for magnitude, from .27 to 1.14 for latency, and from .44 to 1.21 for recruitment.

$(F = 3.98, df = 3/82$, error $MS = .593)$. A similar analysis of amplitude did not result in significant differences $(F = 1.57, df = 3/73$, error $MS = .119)$.

Table 1 summarizes latency and recruitment data on Test Trial 1. For neither recruitment $(F = 1.1, df = 4/90$, error $MS = 1.273)$ nor latency $(F = 1.37, df = 4/90$, error $MS = .775)$ were the differences among groups significant, although it can be noted that the longest latencies occurred in Groups 3 and 5.

*Final nine extinction trials.*—Data of all $S$s who responded at least once in each of three blocks of three extinction trials were employed in the analyses of variance of the four response measures. Table 2 summarizes the means for all groups while Table 3 provides a summary of the analyses of variance. Significant between-groups effects were obtained only with the recruitment and latency measures, in both instances a larger time value being associated with longer ISIs. The significant between-$S$s effect obtained with all measures attests to their reliability.

Both magnitude and latency changed across test trial blocks, with magnitude decreasing and latency increasing. There was a tendency for amplitude to decrease, suggesting that there was some decrease in the size of a response provided that one occurred.

TABLE 3

ANALYSES OF VARIANCE OF MAGNITUDE, AMPLITUDE, LATENCY, AND RECRUITMENT AS A FUNCTION OF ISI OVER THREE BLOCKS OF THREE TEST TRIALS EACH

| Source | df | Amplitude | | Magnitude | | Latency | | Recruitment | |
|---|---|---|---|---|---|---|---|---|---|
| | | MS | F | MS | F | MS | F | MS | F |
| Between $S$s | 80 | | | | | | | | |
| ISI | 4 | .080 | .37 | .572 | .76 | 6.061 | 9.06*** | 8.879 | 6.91*** |
| Error (b) | 76 | .216 | 3.86*** | .752 | 4.82*** | .668 | 2.58*** | 1.285 | 3.65*** |
| Within $S$s | | | | | | | | | |
| Trials | 2 | .143 | 2.56 | 1.716 | 11.00*** | .935 | 3.61* | .521 | 1.48 |
| ISI × Trials | 8 | .056 | 1.00 | .056 | .36 | .264 | 1.02 | .204 | .58 |
| Error (w) | 152 | .056 | | .156 | | .259 | | .352 | |

* $P < .05$.
*** $P < .001$.

TABLE 4

Magnitude Means and SDs for All Ss as a Function of ISI over Three Blocks of Three Test Trials Each

| ISI (Sec.) | N | Test Trial Block | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | |
| | | Mean | SD | Mean | SD | Mean | SD |
| 0 | 20 | 1.411 | .95 | 1.030 | .98 | 1.116 | 1.11 |
| .5 | 22 | 2.260 | .58 | 1.960 | .81 | 1.845 | .95 |
| 1 | 22 | 2.183 | .72 | 1.931 | .92 | 1.915 | .81 |
| 3 | 20 | 1.926 | .89 | 1.892 | .92 | 1.637 | .98 |
| 5 | 22 | 2.007 | .93 | 1.925 | .84 | 1.715 | .99 |

In the above analyses Group 0 had only 9 Ss that met the criterion of at least one response in each block of three test trials. Because of possible bias that might enter by such selection of Ss, a second analysis of magnitude, incorporating data from all Ss, was made. Table 4 provides the means and SDs while Table 5 summarizes the analysis of variance.

Table 5 reveals, first, that there is a significant between-groups effect and, second, that response reliability (as evidenced by the sharply increased F value for between-Ss means) has increased. An examination of mean magnitude in Table 4 indicates that the bulk of the differences between groups rests in Group 0; that is, the increase in N in Group 0 over the prior analysis added sufficient numbers of zeros to provide a between-groups effect.

*Double responses.*—Although preceding analyses were based on the first GSR occurring in the interval from 1 to 5.5 sec. after CS onset, there were, in Groups 3 and 5, frequent instances of a second response with a latency range of from 4 to 8.5 sec. after CS onset. Figure 2, a tracing of one such response from an S in Group 5, illustrates the onset of the first response, a plateau, and then the onset of the second response. In Group 5, of 20 Ss who responded, 19 gave at least one double response. In Group 3, of 18 Ss who responded, 15 gave at least one double response. With the exception of 1 S on one trial in Group 0, no double responses were observed in Groups 0, .5, and 1.

TABLE 5

Analysis of Variance of Magnitude for All Ss as a Function of ISI and Three Blocks of Three Test Trials Each

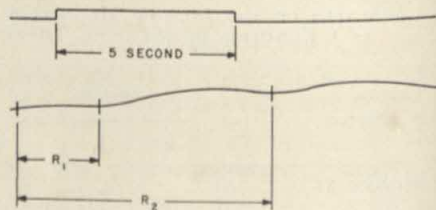| Source | df | MS | F |
|---|---|---|---|
| Between Ss | 105 | | |
| ISI | 4 | 7.502 | 3.71** |
| Error (b) | 101 | 2.022 | 9.96*** |
| Within Ss | 212 | | |
| Trials | 2 | 2.343 | 11.54*** |
| ISI × Trials | 8 | .179 | .88 |
| Error (w) | 202 | 2.03 | |

** P < .01.
*** P < .001.



Fig. 2. A tracing of a double response from 1 S in Group 5. (Onset of the 5-sec. event marker is temporally coincident with the vertical line indicating initiation of $R_1$ and $R_2$, but is displaced due to physical location of event recorder. $R_1$ and $R_2$ are latencies of, respectively, the first and second response.)

## DISCUSSION

The ISI function obtained with magnitude as the dependent variable is similar to that obtained by White and Schlosberg (1952) and Moeller (1954). Within the range tested, results of all three studies suggest that magnitude is greatest near an ISI of .5 sec., similar to the optimum value obtained in numerous studies of skeletal response conditioning (see Kimble, 1961, pp. 156 f.). That amplitude did not vary as a function of ISI suggests that the magnitude function is determined in large part by whether or not $Ss$ in the different ISI conditions responded.[3] On this basis, the distinction between magnitude and amplitude made by Humphreys (1943) and Hilgard (1951) merits further consideration.

Two of the three observations on latency and recruitment have some precedent in the literature: the positive relationship between ISI and recruitment corroborates Bierbaum's (1955) finding and the increase in latency across extinction trials is consistent with Pavlov's (1927, p. 49) report that the latency of salivation increases during extinction trials. We have been unable to find a precedent for the positive relationship between latency and ISI in autonomic conditioning, though such a relationship has been obtained with conditioned skeletal responses (e.g., Boneau, 1958;

[3] Ratios of number of responders to number in each group on Test Trial 1 were 17/20, 22/22, 16/20, and 17/22 for, respectively, Groups 0, .5, 1, 3, and 5. While there is no entirely adequate statistical test available to compare these frequencies for the five groups simultaneously, a Fisher exact test comparing frequency of responders in Group .5 with that of Group 3 and of Group 5 yielded $P$ values of, respectively, .043 and .032 for frequency disparities as large as those obtained. Furthermore, employing the obtained order of mean magnitude (from high to low) in the five groups as a basis for comparison, of the 120 possible ways that the ratios could be ordered only 5 would correspond as well as or better to the magnitude order than that obtained, even if the tied ratio between Groups .5 and 1 were counted as a reversal.

Ebel, 1961). Though data are limited, present evidence suggests that, in addition to magnitude, the very form of the GSR can be brought under the control of external stimulating events. Such a possibility warrants the continued observation of latency and recruitment in GSR research.

Few references have been made to second responses in GSR studies. Rodnick (1937), with ISIs of 17 and 21 sec., found that a second response would occur shortly before shock onset, and Stewart, Stern, Winokur, and Fredman (1961) demonstrated the acquisition of a second GSR while employing an ISI of 7.5 sec. Grings, Lockhart, and Dameron (1961) have shown, further, that this second response is differentiated more rapidly than is the first response in a discrimination learning situation in which an ISI of 5 sec. is employed. While the nature of the second response is not understood, it tends to occur prior to UCS onset. Whether or not its absence from Groups 0, .5, and 1 results because it does not occur in these groups or because recording speed is too slow to separate response rate changes superimposed on the first response remains to be investigated.

## SUMMARY

Five groups of $Ss$ trained at CS-UCS intervals of 0, .5, 1, 3, or 5 sec. were employed in a study of the role of the interstimulus interval (ISI) in the conditioning of the GSR. All $Ss$ received 20 tone-shock pairings followed by 10 tone alone (test) trials. Four attributes of the GSR were measured: amplitude, magnitude, latency, and recruitment.

The principal findings were: (a) magnitude was greatest with an ISI of .5; (b) magnitude, but not amplitude, varied with ISI; (c) latency and recruitment both increased as a function of ISI; (d) latency increased across extinction trials; and (e) a second response was observed frequently in $Ss$ exposed to ISIs of 3 and 5 sec.

## REFERENCES

BIERBAUM, W. B. Temporal aspects in conditioning of the GSR. Unpublished doctoral dissertation, University of Florida, 1955.

Boneau, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

Ebel, H. C. Stable-state behavior and reversibility as a function of the inter-stimulus interval variable in classical eyelid conditioning. Unpublished master's thesis, Pennsylvania State University, 1961.

Grings, W. W., Lockhart, R. A., & Dameron, L. E. Interstimulus interval as a variable in GSR conditioning of mentally deficient individuals. Paper read at American Psychological Association, New York, 1961.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

Humphreys, L. G. Measures of strength of conditioned eyelid responses. *J. gen. Psychol.*, 1943, **29**, 101–111.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* (2nd. ed.) New York: Appleton-Century-Crofts, 1961.

Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

Pavlov, I. *Conditioned reflexes.* New York: Oxford Univer. Press, 1927.

Rodnick, E. H. Characteristics of delayed and trace-conditioned responses. *J. exp. Psychol.*, 1937, **20**, 409–425.

Stewart, M. A., Stern, J. A., Winokur, G., & Fredman, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60–67.

White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

# ON-TARGET VERSUS OFF-TARGET INFORMATION AND THE ACQUISITION OF TRACKING SKILL [1]

ALTON C. WILLIAMS [2] AND GEORGE E. BRIGGS

*Ohio State University*

Knowledge of results or information feedback has been an area of substantial research interest since the classic study of Thorndike (1927), the most recent review for motor skills being that of Bilodeau and Bilodeau (1961). The present study was concerned with a restricted topic within the more general area of knowledge of results: the influence of *augmented feedback* on the acquisition of skilled performance. As the term implies, augmented feedback represents knowledge of performance in addition to that normally present in a skill task and it is defined as secondary information supplemental to some primary feedback signal(s).

Of the numerous studies on acquisition of tracking skill only those by Rapparlie (see Bray, 1948, pp. 195–196) and by Payne and Hauty (1955) have employed augmented feedback based on an off-target criterion, the majority of *E*s having activated such signals when tracking error was within tolerance, i.e., when *S* was on-target. Thus, augmented feedback has been utilized in a majority of the research as a means of emphasizing "correct" responding, a use which seems logical

in view of the Skinnerian position that errors or incorrect responses should be minimized in the construction of training tasks and programs, at least for discrete-verbal skills. However, it is an open question whether in continuous motor skill tasks *error* information feedback might not be more effective than information on correct responding, and the present study was concerned, therefore, with the effects on tracking performance of augmented feedback based on an off-target criterion relative to that based on an on-target criterion. Further, a simple off-target criterion condition and an off-target condition providing directional information were compared.

## METHOD

*Apparatus.*—The skill task was defined by the SETA apparatus (Gain & Fitts, 1959) which provides *S* with a simple positional tracking task. The *S* tracked a 6 cpm + 12 cpm sinusoidal signal via a one-dimensional compensatory display of tracking information on a 5-in. cathode ray tube. This display was noise free and provided *S* with his primary source of feedback information. Augmented feedback was provided to the experimental groups in the form of auditory clicks delivered to headphones worn by *S*. The clicks occurred at the rate of two per sec. when *S* was within or outside (depending upon the experimental condition) preset tolerance limits of system error. The tolerance limits were attained by adjusting two voltage comparators to the desired voltage levels such that a plate voltage relay was activated whenever tracking error fell within the critical levels. Activation of the relay either closed a circuit from the click generator to the headphones or opened that circuit depending upon the experimental condition. Further, it was possible to activate only the left earphone when tracking error was to the left on

the visual display and only the right earphone in the opposite condition. The on-target band thus defined represented tracking error of rather small amplitudes: in terms of the visual display, the on-target band was ±.08 in. around the fixed target element.

Performance was scored by electronically integrating the absolute value of tracking error. This is the analog of the average deviation of S's error amplitude distribution and will be identified as average error. The scores were transformed from voltage units to units of inches on the display scale; thus, average error as reported here may be interpreted as the average deviation of S's tracking error amplitude distribution plotted on the same scale used in the visual tracking display.

In addition to average error, E recorded the amount of time during each tracking trial when tracking error was within the tolerance limits. These data were transformed to percent time within tolerance scores.

*Subjects and procedure.*—The Ss were 88 volunteer male undergraduates. Each S participated in five daily 30-min. sessions, and none of the Ss had previous experience with a laboratory tracking task.

Twenty-two Ss were assigned via a randomization procedure to each of four groups. The control group, Group C, did not experience augmented feedback at any time during the experiment. The three experimental groups experienced the auditory clicks during the initial three and one-half training sessions whereupon augmented feedback was withdrawn for the remaining one and one-half sessions. Group I received augmented feedback whenever tracking error was *within* the preset tolerance limits (see above), while Groups O and O-D received the signals whenever tracking error exceeded the on-target limits. For Groups I and O augmented feedback occurred simultaneously in both earphones, while for Group O-D the left earphone was activated when displayed error was to the left of the fixed target reference on the visual display, and the right earphone was activated when error was to the right.

All tracking trials were of 30-sec. duration and were administered in blocks of four trials. There were 3 blocks of trials on the first and last sessions and 4 blocks in each of the other sessions, a total of 18 blocks. The initial 13 blocks defined the training conditions for the experimental groups, while transfer to the no-augmented-feedback condition occurred over the final 5 blocks of trials. Rest periods of 30 sec. were provided between trials within a block and 1.5 min. rest occurred between blocks. Performance was scored over the final 25 sec. of each trial, which permitted the initial transients in tracking behavior to dampen out prior to scoring.

The instructions to all Ss stressed the importance of maintaining zero error on the visual display, and each S was reminded of this ultimate goal at the beginning of each session.

## RESULTS

*Average error.*—Figure 1 provides a summary of tracking accuracy as defined by the average error scores. It may be noted that all three experimental groups attained proficiency levels superior to that of Group C. Further, Group O was superior to both Groups I and O-D throughout both training and transfer, and there was considerable overlap between the latter two groups during training. It appears, then, that augmented feedback based on a simple off-target criterion (Group O) resulted in performance superior to that attained either with an on-target criterion (Group I) or with an off-target criterion which included directional information (Group O-D).

These observations are supported by the results of the Mann-Whitney U test which was applied to the average error data for groups. The tracking error scores were summed for each S across all 18 trial blocks for this analysis; thus, $n_1 = n_2 = 22$ for each two-group comparison performed. The alternative hypothesis tested here was that more Ss in one of a pair of groups were superior in tracking proficiency than could occur by chance. A nonparametric analysis was employed since marked heterogeneity of variance was found in the data (see Fig. 2). The results of the U tests are listed in the upper half of Table 1, where it may be noted that all group comparisons accepted the alternative hypothesis at $P < .05$ except that between Groups I and O-D.
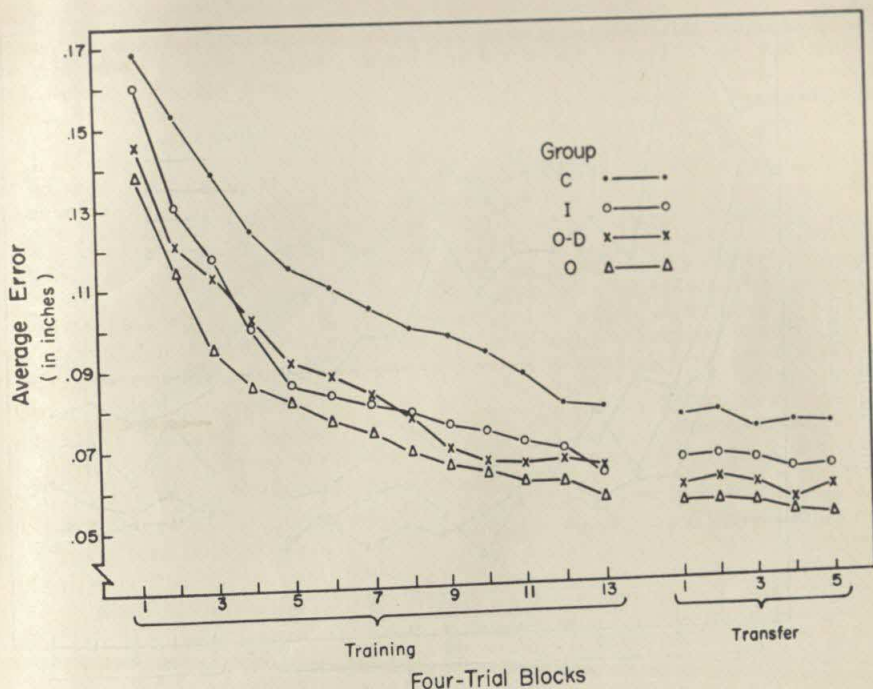
Fig. 1. Average tracking error for all groups during training and transfer.

*Variability.*—Augmented feedback also exerted an influence on inter-$S$ variability within groups (intragroup variability), and these data are summarized in Fig. 2. The data plotted are intragroup $SD$s of the average error scores for each block of training and transfer trials. As with the average error data (Fig. 1), the three experimental groups are superior to the control group in terms of intragroup variability. However, here, Group O does not enjoy a clear superiority over Group O-D, as was the case in Fig. 1, but both Groups O and O-D do exhibit more inter-$S$ homogeneity than that attained by Group I.

These observations are supported by the results of $U$ tests which were applied to each pairing of groups. It was not possible to test the same hypothesis as in the case of the average error data; instead, the alter-native hypothesis tested with the intragroup $SD$s was that more intragroup $SD$s for one of a pair of groups would be smaller in value than those for the other group than could occur by chance. For each comparison $n_1 = n_2 = 18$. The results are listed in the lower half of Table 1 where it may be seen that (a) Groups O and O-D exhibited more intragroup homogeneity than did Group I, and (b) all three experimental groups were superior in this regard to Group C. It may be concluded, then, that augmented feedback not only results in a higher level of performance accuracy but also individual differences within groups are less than is the case for $S$s who do not experience these additional feedback cues.

*Time within tolerance.*—The percentage of time $S$ spent within the $\pm.08$-in. tolerance limits (TWT) is
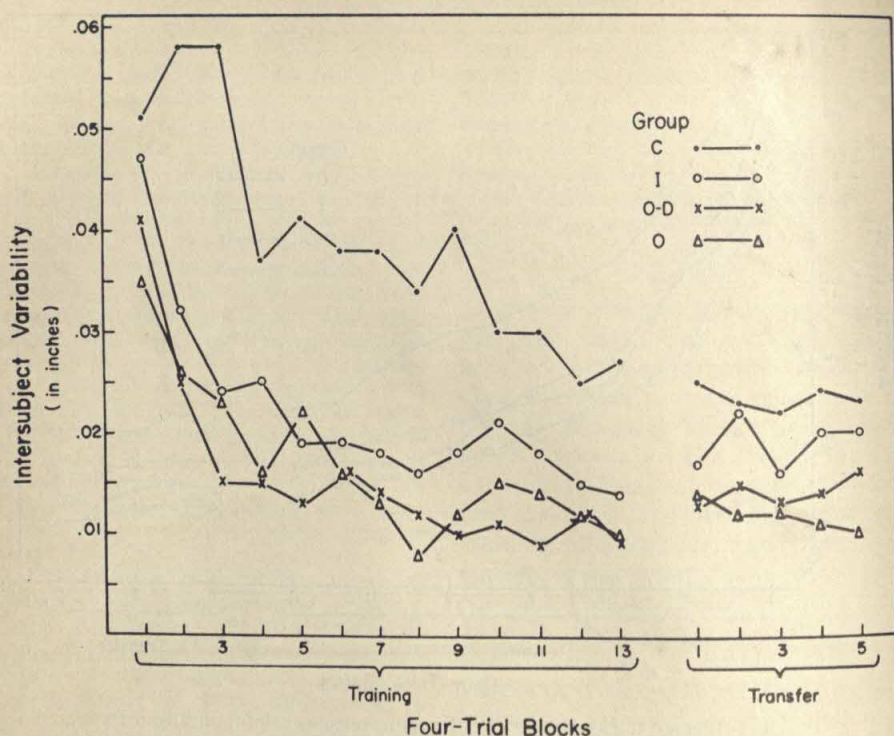
FIG. 2.    Inter-S variability for all groups during training and transfer.

summarized for each four-trial block in Fig. 3. While the differences among experimental groups are not as apparent as with the average error data, there are statistically significant ($P < .01$) differences both during training and transfer (see Table 2).

The pattern of group proficiency levels on TWT differs in an important way from that found with the average error data: in terms of the TWT scores Group I is significantly superior to all other groups during training at $P < .05$. This rank order for Groups

TABLE 1

RESULTS OF THE MANN–WHITNEY $U$ TEST APPLIED TO AVERAGE
ERROR AHD INTER–$S$ VARIABILITY DATA

| Performance Criterion | | Groups Compared [*] | | | | | |
|---|---|---|---|---|---|---|---|
| | | C/I | C/O-D | C/O | I/O-D | I/O | O-D/O |
| Average Error ($n_1 = n_2 = 22$) | $U$ | 333 | 346 | 394 | 270 | 330 | 339 |
| | $z$ | 2.14 | 2.44 | 3.57 | 0.66 | 2.07 | 2.28 |
| | $P$ | .033 | .014 | .001 | .510 | .038 | .022 |
| Variability ($n_1 = n_2 = 18$) | $U$ | 33 | 19 | 16 | 46 | 63 | 149 |
| | $P$ | <.002 | <.002 | <.002 | <.002 | <.002 | >.10 |

[*] For average error: less/more accurate group; for variability: more/less variable group.
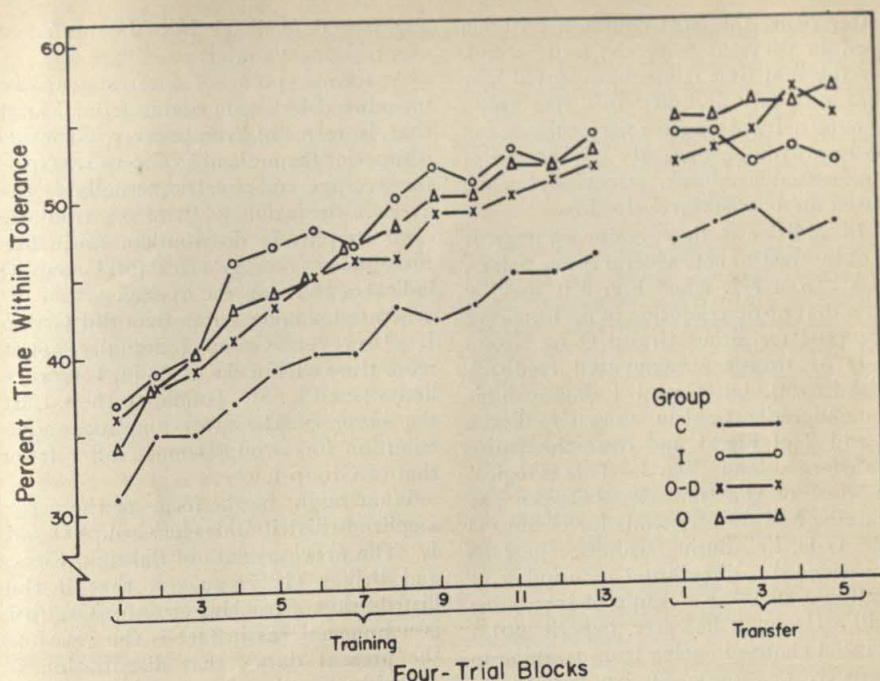
FIG. 3. Percent time within tolerance for all groups during training and transfer.

I and O is just the reverse of that noted for the average error measure where Group O was superior, and this

## TABLE 2
ANALYSIS OF VARIANCE OF PERCENT TIME WITHIN TOLERANCE DATA FOR TRAINING AND TRANSFER

| Source | df | MS | F |
|---|---|---|---|
| Training | | | |
| Groups (G) | 3 | 2741.13 | 7.22** |
| Ss within groups (Ss/G) | 84 | 379.54 | |
| Blocks (B) | 12 | 2470.55 | 182.66*** |
| B × G | 36 | 20.02 | 1.48* |
| B × Ss/G | 1008 | 13.52 | |
| | | | |
| Transfer | | | |
| Groups (G) | 3 | 1314.14 | 5.34** |
| Ss within groups (Ss/G) | 84 | 246.16 | |
| Blocks (B) | 4 | 17.18 | 1.21 |
| B × G | 12 | 39.41 | 2.78** |
| B × Ss/G | 336 | 14.20 | |

\* $P < .05$.
\*\* $P < .01$.
\*\*\* $P < .001$.

reversal serves to define how Ss of Groups O and I responded differentially to the task (see below).

Finally, it may be noted that Group I actually deteriorated during transfer while Groups C, O, and O-D either "held their own" or improved in terms of TWT. This deterioration gave rise to the significant Blocks × Groups interaction in the transfer analysis of Table 2.

## DISCUSSION

The majority of past research in this area has found superior performance by groups trained with augmented feedback compared to no-augmented-feedback control groups. In this regard, the present research merely confirms those previous data. The contribution of the present study is twofold: first, the use of an off-target criterion for the activation of augmented feedback was found to be superior to an on-target criterion, the

latter being the most common criterion used in previous research; and second, for the first time it has been noted here that not only accuracy but also group homogeneity of tracking performance can be improved significantly by the use of augmented feedback, especially when based on an off-target criterion.

In addition to these major findings, it is of interest to note several other points. First, from Fig. 1 and Fig. 3 it may be seen that no deterioration in performance occurred for either Group O or Group O-D at transfer (augmented feedback withdrawn), but Group I does exhibit some deterioration on Transfer Blocks 1 and 2 of Fig. 1 and over the entire transfer session in Fig. 3. This is logical in view of the similarity between the training and transfer trials for Groups O and O-D, i.e., during training these $S$s experienced a diminution in number of auditory clicks as a result of increasing skill. Group I, however, experienced a marked change in going from training to transfer, i.e., these $S$s experienced increasing amounts of auditory clicks with training followed by an abrupt change to no-augmented-feedback at transfer. These observations were supported by statistical analyses of the transfer data: Groups O and O-D were superior to Group C throughout transfer in terms of tracking accuracy ($P < .05$), but Groups I and C did not differ ($P > .05$) from the first transfer trial to the last.

The reason for the loss of superiority by Group I over Group C may be found in Fig. 2: upon transfer Group I exhibited an increase in inter-$S$ variability considerably greater than the modest increase shown by Groups O and O-D. It follows, then, that the transfer performance deterioration by Group I involved not only a loss in tracking accuracy (Fig. 1) but also an increase of within-group variability (Fig. 2). Since apparent accuracy deterioration is not noted for Groups O and O-D, and since the increase in inter-$S$ variability was relatively less than that for Group I during transfer, it follows that an off-target criterion for the activation of augmented feedback is to be preferred as a training variable.

A second point of interest can be introduced by again noting from Fig. 1 that, in terms of average error, Group O is superior throughout to Group I. Now, the average error metric actually is the average deviation ($AD$) of $S$'s tracking error amplitude distribution, and thus the smaller average error for Group O indicates that, on the average, those $S$s generated smaller errors than did Group I. However, Group I actually spent more time within the $\pm.08$-in. tolerance limits (see Fig. 3). It follows, then, that the *shape* of the error amplitude distribution for Group O must differ from that of Group I.

What might be the forms of the error amplitude distributions for Groups O and I? The previous data of Bahrick, Fitts, and Briggs (1957) suggest that if the distribution of tracking error for Group I is nonnormal (as in fact is the case for the present data), that distribution is probably leptokurtic. On this basis, it follows that Group O generated a *bimodal* error amplitude distribution, as this is the only distribution that could result in both smaller $AD$ and smaller percent time within tolerance scores for Group O compared to Group I. In other words, $S$s of Group O spent considerable time tracking closely *around* the two tolerance limit points and made relatively few errors of large amplitude, while Group I spent more time *within* the tolerance limits but committed occasional large tracking errors.

These deductions from the data, then, lead one to conclude that tracking behavior with augmented feedback based on an off-target criterion differs fundamentally from that when such feedback is based on an on-target criterion: in the former case large errors are emphasized, and $S$ apparently learns rather quickly to minimize these occasional lapses in tracking accuracy, while in the latter case the importance of small errors is emphasized, but apparently $S$ does not respond as quickly or as efficiently to correct for occasional large tracking errors. Since minimizing large tracking

errors is a primary task for $S$ early in training, it follows that augmented feedback based on an off-target criterion should be particularly helpful in shaping the desired behavior during early acquisition of a continuous control skill. It is problematical whether or not the above generalization will hold in discrete-verbal learning tasks. However, it is suggested that if the element of discovery is present in such discrete tasks, some criterion for the activation of augmented feedback analogous to the off-target criterion in the continuous case would be helpful in focusing $S$'s attention on response alternatives reasonably close to the correct alternative.

Finally, it is interesting to note that the additional feedback information available to Group O-D (off-target direction of error) did not provide for a performance level superior to that of Group O which received augmented feedback on a more simple, nondirectional off-target criterion.

Two explanations for the relative inferiority of Group O-D are suggested. First, it is possible, of course, that the addition of a directional cue to the off-target criterion was not particularly useful information to Group O-D since directional relationships between control and display movements are one of the most simple aspects of a tracking task to be learned. Secondly, it is possible that the information on error direction was actually disruptive. It may be recalled that auditory clicks were delivered to the left earphone when error was to the left on the visual tracking display. Thus, a signal in the *left* earphone indicated that $S$ should move his control *to the right*. While data are lacking on population stereotypes in such a stimulus-response task, it is probable that a more compatible S-R arrangement would be one in which a signal in, say, the *left* earphone indicates a control movement is required *to the left*. It is suggested that both of the above possibilities were responsible for the inferior performance of Group O-D relative to that of Group O.

## SUMMARY

During training three experimental groups received augmented feedback (auditory clicks at the rate of two per sec.) when tracking accuracy was within (an on-target criterion) or outside (an off-target criterion) fixed tolerance limits. During transfer, no augmented feedback was provided. All three experimental groups were superior in tracking accuracy during training to a control group which did not receive augmented feedback. Of the three experimental groups, the group receiving augmented feedback when off-target was superior to a group which experienced clicks when on-target. It was superior also to a group which received clicks when off-target but differentially according to the direction of tracking error.

During transfer both off-target groups remained superior in tracking accuracy to the control group, but the on-target group and the control group attained comparable performance.

It follows that augmented feedback based on a simple off-target criterion was the most effective training condition. An analysis of the data suggested that this superiority was a result of the emphasis an off-target criterion places on occasional large tracking errors. The group trained on this condition apparently learned to reduce such errors more quickly and efficiently than did the on-target criterion group.

## REFERENCES

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. Learning curves: Facts or artifacts? *Psychol. Bull.*, 1957, **54**, 256–268.

Bilodeau, E. A., & Bilodeau, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

Bray, C. W. *Psychology and military proficiency.* Princeton, N. J.: Princeton Univer. Press, 1948.

Gain, P., & Fitts, P. M. A simplified electronic tracking apparatus (SETA). *USAF WADC tech. Rep.*, 1959, No. 59-44.

Payne, R. B., & Hauty, G. T. Effect of psychological feedback upon work decrement. *J. exp. Psychol.*, 1955, **50**, 343–351.

Thorndike, E. L. The law of effect. *Amer. J. Psychol.*, 1927, **39**, 212–222.

# RESISTANCE TO EXTINCTION AFTER VARYING AMOUNTS OF DISCRIMINATIVE OR NONDISCRIMINATIVE INSTRUMENTAL TRAINING [1]

## M. R. D'AMATO, DONALD SCHIFF, AND HARRY JAGODA

*New York University*

Resistance to extinction and number of reinforced responses (acquisition level) have classically been thought to be monotonically related. Recently, however, doubt concerning the monotonicity of the relationship has been raised in several quarters (e.g., Birch, Ison, & Sperling, 1960; Murillo & Capaldi, 1961; Senko, Champ, & Capaldi, 1961). Although nonmonotonicity between resistance to extinction and acquisition level has been reported in a rather large number of recent studies, many of these are of questionable relevance because either they made use of some type of an intermittent reinforcement schedule (e.g., Capaldi, 1957, 1958) or they employed a discriminative rather than a nondiscriminative task (Murillo & Capaldi, 1961), in some cases "extinction" actually constituting reversal learning (e.g., Senko et al., 1961). In certain other studies reporting nonmonotonicity and not subject to either of the preceding objections (Lewis & Duncan, 1956, 1958), one finds other reasons for questioning their pertinence for the present problem, such as the use of human *S*s in money payoff situations, which bear only the slightest resemblance to the strictly instrumental situations in which, with animal *S*s, the relationship of monotonicity was originally established.

Of the several published studies falling in the latter category, two have reported resistance to extinction and acquisition level to be nonmonotonically related. North and Stimmel (1960) found that extinction of the running response proceeded more rapidly for *S*s given 90 or 135 rewarded trials in a runway than for *S*s given 45 reinforced trials. Wilson (1958), on the other hand, found no evidence of a reduction in resistance to extinction, even though he carried (runway) acquisition to 480 trials. Harris and Nygaard (1961), working with thirsty rats in a free operant situation, observed the usual monotonic relationship between acquisition level and resistance to extinction of the bar pressing response up to 360 reinforced responses, the highest acquisition level they worked with. Under rather similar conditions, Margulies (1961) reported monotonicity up to 1,000 trials; and as described in an earlier paper (D'Amato & Jagoda, 1962), unpublished results from our laboratory suggest persistence of the monotonic relationship up to some 7,000 rewarded bar pressing responses, again with thirsty rats as *S*s. Finally, in conflict with the preceding results, King, Wood, and Butcher (1961) recently reported a reduction in the resistance to extinction of pigeons receiving 600 or 900 reinforced key pecks, as compared to *S*s permitted 300 rewarded responses.

In summary, the results from studies employing a nondiscriminative (simple) instrumental response are equivocal, though they suggest that under certain conditions the monotonicity assumption holds up to several thousands of responses.

The present study had three objectives. First, all of the cited Skinner box studies in which monotonicity was obtained employed the thirst drive and water reward. It would be of some value to demonstrate that monotonicity also holds in the Skinner box

setting for the hunger drive and food reward, particularly since all other relevant studies have been conducted under the latter motive-incentive conditions. Second, we wished to carry acquisition training beyond the level achieved in previously published reports, to a maximum of 1,600 reinforcements. Third, and most important, we wished to assess the influence of two different training procedures, discriminative versus non-discriminative instrumental training, on the shape of the function relating acquisition level and resistance to extinction. The hypothesis under examination was that the traditional monotonic relationship would prevail for Ss given nondiscriminative training, but not for Ss given discrimination training, the latter being expected to show a significant decline in resistance to extinction with prolonged acquisition training.

One plausible rationale for the preceding hypothesis is as follows. In general, Ss trained on a discrimination program are essentially on an intermittent reinforcement schedule (with respect to the experimental situation) until the discrimination is firmly acquired; beyond this point their experience is to all intents and purposes one of continuous reinforcement (since they then make few errors, or respond little in S$^\Delta$). Prolonged discrimination training, then, has the effect of carrying Ss beyond their intermittent reinforcement experience well into the continuous reinforcement segment. If it is assumed that directly following an intermittent reinforcement experience resistance to extinction is greater than when a prolonged continuous reinforcement segment intervenes between the intermittent reinforcement experience and extinction, the hypothesis follows.

## METHOD

*Subjects.*—The Ss that completed the study were 94 experimentally naive albino rats (51 males and 43 females) 53 to 94 days of age at the start of the study. One S was discarded because of a programing error and a second due to illness. All Ss were bred in our laboratory.

*Apparatus.*—Two Grason-Stadler two-bar Skinner boxes were used; the right bars were removed from both boxes, converting them into single-bar boxes. A force of approximately 20 gm. was required to activate the bar microswitches in the two boxes.

*Design.*—The experiment was run in two replications (of unequal N), but since the results of the two replications were quite similar in form, replications was not included as a factor in the statistical design.

There were four levels of acquisition, defined in terms of the number of reinforced responses allowed after pretraining: 200, 400, 800, and 1,600. At each acquisition level there were two separate groups of Ss that differed in the type of training received. The "I" groups, designated as I(200), I(400), I(800), and I(1600), underwent simple (nondiscriminative) instrumental training, each of the four groups being composed of 12 Ss. The "D" groups, D(200), D(400), D(800), and D(1600), were trained on a successive brightness discrimination problem. There were 11 Ss in the first two groups and 12 each in the last two groups.

All Ss of a replication were quasirandomly assigned to the eight groups, in part balancing litters over groups, and placed on deprivation at the same time. The I and D groups of a given acquisition level started acquisition training together; the scheduling of the beginning of acquisition was so arranged for the various reinforcement groups that all groups entered extinction within a day or two of each other.

*Deprivation training.*—One week prior to the beginning of pretraining, Ss were placed on a 22-hr. food deprivation training regimen, water being constantly available in the home cages. Four days later, the feeding period was reduced to 1½ hr. daily and remained at that duration throughout the study.

*Skinner box pretraining.*—There were 3 pretraining days. On Days 1 and 2, 50 reinforcements (45-mg. Noyes rat pellets) were given in the conditioning of approach responses to the food tray at the sound of the feeder magazine. On Day 3, Ss were shaped on the bar pressing response with 25 to 50 reinforcements. The stimulus conditions in

the boxes during pretraining were the same as prevailed during simple instrumental training and during the $S^D$ portion of discrimination training.

*Discrimination training.*—All D groups were exposed to a simple brightness discrimination situation. Illumination of the left (white jeweled) stimulus light constituted $S^D$ and provided an illumination level of 7 to 10 ft-c, measured with the target of a Weston illumination meter placed at bar level and $\frac{3}{4}$ in. in front of the lens of the stimulus light. Measured under the same conditions the illumination under $S^\Delta$, which was produced by the shielded house light, read about 0.1 ft-c.

All $S^D$ periods were 45 sec. in duration and alternated with $S^\Delta$ periods varying between 33 and 65 sec. in length. One revolution of the programing film tape took about 6 min. and provided approximately equal total times in $S^D$ and $S^\Delta$. Since 50 reinforced responses in $S^D$ were allowed each day, Groups D(200), D(400), D(800), and D(1600) required 4, 8, 16, and 32 training days, respectively.

*Simple instrumental training.*—The treatment of the I groups was exactly the same as that accorded the corresponding D groups except for the elimination of the $S^\Delta$ periods, i.e., the left stimulus light was always on. As in the D groups, 50 reinforcements were permitted on each acquisition day.

At the end of the daily 50 reinforced responses, Ss of all groups entered a time-out period and were quickly removed from the Skinner boxes to their home cages where, no sooner than 15 min. later, they were fed for the 1½-hr. period. The estimated average number of hours of food deprivation at the start of a day's session was 22 hr.

*Extinction.*—The extinction procedure, which began the day following the termination of acquisition, was precisely the same for all Ss and consisted of one daily 10-min. period on each of 5 successive days with the left stimulus light illuminated, i.e., in the former $S^D$ condition. The duration of the extinction periods, 10 min., was judged to be short enough to resemble closely the length of the acquisition sessions (which overall averaged 7.15 min.) and yet long enough to sample adequately the extinction process. It should be pointed out, however, that because of the absence of $S^\Delta$ periods, the extinction sessions bore a greater similarity to the acquisition sessions of simple instrumental training than to those of discrimination training. The exclusion from extinction of $S^\Delta$ periods (more correctly, periods in which the former $S^\Delta$

stimulus was present) was based on two considerations: (a) We wished to make direct comparisons of the extinction performance of corresponding I and D groups, which would be feasible only if the extinction sessions were identical for all Ss. (b) Inclusion of such periods would have raised problems concerning the treatment of responses made in $S^\Delta$, since the number of such responses most probably would be related to the acquisition level variable.

It should be recorded that the cue provided by activation of the feeder magazine was maintained during extinction. Deprivation level at the start of extinction sessions was confined within the limits of 21 to 23 hr. During the course of the experiment the relative humidity ranged between 58% and 70%, and the temperature between 70° and 78° F.

## RESULTS

*Discrimination acquisition.*—A discrimination ratio (DR), obtained by dividing the number of responses made in $S^\Delta$ by the number performed in $S^D$, was calculated for each S after every daily training session. In a plot of the daily means of the DRs, Group D(1600) showed some improvement over the level of discrimination finally achieved by Group D(800). To evaluate this difference the mean DR over the last 8 acquisition days was calculated for every S of Group D(1600), as was the mean DR over the last two sessions for each S of Group D(800). The group means based on these measures (.141 and .218, respectively) did not, however, differ significantly ($t = 1.67$, $df = 22$, $P \approx .12$).

*Extinction in the I groups.*—The number of responses made by each S in each of the five daily extinction sessions was converted to common logs, providing the basis for all statistical analyses. An overall index of resistance to extinction was obtained for each S by summing its five daily log scores.

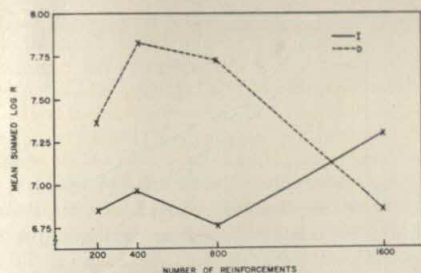The first question of interest concerns the relationship observed be-

Fig. 1. Relationship between acquisition level (number of reinforcements) and resistance to extinction in the simple instrumental (I) and the discrimination (D) groups.

tween resistance to extinction and the number of reinforced responses allowed during acquisition. The solid line in Fig. 1, based on the means of the summed log scores, presents this relationship. It is plain from the figure that there is no tendency for resistance to extinction to decrease with increasing training, even when acquisition is carried to 1,600 reinforced responses; in fact, the solid line suggests an increasing trend.

The most powerful (and specific) way of analyzing the present data is by a trend analysis of the linear and quadratic components of the curve (Grant, 1956). The presence of a significant quadratic component (with negative sign) alone or in accompaniment with a significant negative linear component would constitute evidence for a nonmonotonic relationship between acquisition level and resistance to extinction. A significant negative linear component in conjunction with an insignificant quadratic would, essentially, support the same interpretation. On the other hand, an insignificant quadratic component would support the assumption of monotonicity if the linear component were either insignificant or positive in sign. The latter would indicate that the curve was still rising, while the occurrence of insignificance

in both the linear and quadratic components would signify that the function was asymptotic.

In order to take into consideration the unequal spacing of the independent variable, the coefficients of the orthogonal polynomials were calculated in the manner suggested by Grandage (1958). Analysis of variance showed both the linear and the quadratic components of the trend to be insignificant ($F = 1.29$ and $0.59$, respectively, $df = 1/44$). Thus, for the numbers of rewarded responses employed in this study, the function relating resistance to extinction and acquisition level in the nondiscriminatively trained $S$s was essentially asymptotic.

Despite the absence of significant differences in the summed log extinction scores, one would like to know whether extinction in the four I groups followed a parallel development over the five extinction sessions. Figure 2 presents the extinction curves of the four I groups over the 5 extinction days. A trend analysis
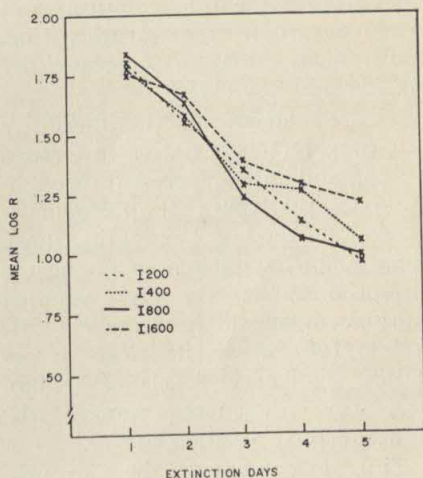


Fig. 2. Extinction curves of the four simple instrumental groups, based on the log of the number of daily extinction responses.

(Edwards, 1960) was applied to the data and the differences among the groups' linear components evaluated. The appropriate $F$ (2.38, $df = 3/176$) fell short of accepted significance levels ($P \approx .07$). The curves of Fig. 2 suggest that more impressive differences among the linear components of the groups' trends, as well as among their overall extinction scores, might have been obtained if extinction had been carried one or two sessions further. At any rate, the data of the figure plainly show that Group I (1600) is not inferior to any of the other I groups.

*Extinction in the D groups.*—The first concern is again with the relationship between acquisition level and resistance to extinction, which, as may be seen from Fig. 1, is vastly different from that obtained with the I groups. A trend analysis of the curve revealed a significant linear component ($F = 4.13$, $df = 1/42$, $P < .05$), as well as a significant quadratic component ($F = 6.31$, $df = 1/42$, $P < .025$). Because of the unequal $N$s in the four discrimination groups, the trend analysis was based on groups means, rather than sums, in the manner suggested by Walker and Lev (1953).

Figure 3 shows that the inferiority of Group D (1600) relative to Groups D (400) and D (800) is present throughout extinction, though it appears most marked on the last 2 extinction days. Once again the differences among the groups' linear trends were evaluated and once again the resulting $F$ (2.51, $df = 3/168$) was quite close to accepted significance levels ($P \approx .06$), suggesting differences among the slopes of the extinction curves.

Thus, in contrast to the I groups, the discriminatively trained $S$s, after an initial increase, reveal a sharp reduction in resistance to extinction
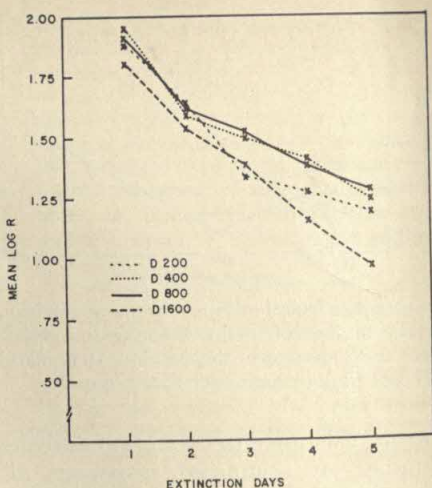


FIG. 3. Extinction curves of the four discrimination groups, based on the log of the number of daily extinction responses.

as a function of increasing acquisition training. Further, the nonmonotonicity appears in the early stages of extinction, tending to be somewhat more marked during the latter extinction days.

*Comparison of extinction in the I and D groups.*—If it is in fact true that discrimination training, at least in its early phases, provides an intermittent reinforcement experience, then resistance to extinction should be greater in the combined D groups than in the I groups. Analysis of the differences between the means of the summed log scores of the combined four D groups (7.43) and the four I groups (6.96) revealed that discrimination training did indeed lead to significantly greater resistance to extinction ($t = 2.53$, $df = 86$, $P < .02$).

It will be observed in Fig. 1 that a reversal in the extinction curves of the I and D groups occurs in the 1600 groups; the reversal is, however, far from significant ($t = 1.25$, $P > .20$). Finally, tests of differences between the means of the I and D groups receiving 200, 400, and 800 reinforce-

ments yielded, in the same order, the following results: $t = 1.45$, $P > .10$; $t = 2.25$, $P < .05$; $t = 2.63$, $P < .02$.

## DISCUSSION

In general, our results support the hypothesis that the presence or absence of monotonicity in the function relating acquisition level to resistance to extinction depends importantly on the type of learned response under consideration. In agreement with earlier reports (Harris & Nygaard, 1961; Margulies, 1961) monotonicity was found with a nondiscriminative (free operant) response; with a discriminative response, on the other hand, a marked nonmonotonicity between acquisition level and resistance to extinction was observed. While the present results do little to explain the nonmonotonicity obtained in the presumably nondiscriminative situations of North and Stimmel (1960) and King, Wood, and Butcher (1961), they perhaps shed some light on other, related, studies in which a strong discriminative component was present (e.g., Murillo & Capaldi, 1961; Senko et al., 1961). Conceivably, they also possess some relevance for the overlearning reversal effect (the faster reversal learning of Ss receiving extensive overtraining), inasmuch as that phenomenon has been attributed by some to a nonmonotonic relationship between acquisition level and resistance to extinction of the approach response to the originally positive stimulus (e.g., Birch et al., 1960).

Turning now to a possible mechanism by which the nonmonotonicity of the discrimination groups might be explained, two separate factors seem to be involved. First, it probably is safe to assume that the intermittent reinforcement experience unavoidably associated with the early phases of discrimination training has the effect of augmenting resistance to extinction in Ss receiving moderate amounts of discrimination training. This assumption is supported in the present study by the superior resistance to extinction of the combined discrimination groups, with further veri-

fication coming from an earlier study by Jenkins (1961a), who worked with pigeons in a discrete trials situation.

Second, it is still necessary to explain how resistance to extinction becomes depressed with overtraining, and two possibilities suggest themselves. The first one is that mentioned earlier, namely, that overtraining has the effect of carrying Ss beyond their intermittent reinforcement experience well into a region of virtual continuous reinforcement. However, there are the following difficulties with this possibility. (a) Several studies (e.g., Jenkins, 1961b; Theios, 1962) have failed to demonstrate a reduction in resistance to extinction as a result of interpolating a continuous reinforcement segment between a partial reinforcement experience and subsequent extinction. (b) There is evidence that nonmonotonicity between acquisition level and resistance to extinction can occur in situations where responding to the negative stimulus is under the control of E rather than S, as in successive discrimination training on a straightaway (Birch et al., 1960). (c) Finally, the nature of the present argument is such that, in principle, no amount of discrimination training could reduce resistance to extinction below that of a comparable nondiscriminative group. The results obtained with the 1,600 groups of the present study suggest that such a reversal is a distinct possibility if acquisition were carried somewhat further.

An alternative interpretation of the basis of the nonmonotonicity maintains that the discrimination experience is vital completely apart from attending changes in the effective reinforcement schedules (cf. Murillo & Capaldi, 1961, who found nonmonotonicity only in those Ss that had learned the discrimination presented them during acquisition). This is a position, however, that is difficult to specify with any degree of precision. Nevertheless, it is suggestive that in Group D(1600) goodness of discrimination (the reciprocal of the mean of the DRs over the last 10 acquisition days) was negatively cor-

related with resistance to extinction (summed log R), with rho equal to .53 (.05 < P < .10).

## SUMMARY

This experiment investigated the hypothesis that acquisition level and resistance to extinction would be monotonically related for a simple (nondiscriminative) instrumental response (bar pressing in a Skinner box), but the function would be nonmonotonic for a comparable discriminative response (successive brightness discrimination). Four groups of Ss were trained on the simple instrumental response and allowed 200, 400, 800, or 1,600 reinforced responses, 50 per day. The same numbers of reinforced responses were given to four corresponding groups of discriminatively trained Ss, the procedure employed with the latter differing only in the insertion of occasional $S^{\Delta}$ periods. All groups were exposed to one 10-min. extinction period (in the former $S^{\Delta}$) on each of 5 successive days. Trend analyses of the data supported the initiating hypothesis; and as expected, the discriminatively trained Ss were, as a group, more resistant to extinction than the Ss trained on the simple instrumental response.

## REFERENCES

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, 60, 36–40.

CAPALDI, E. J. The effects of different amounts of alternating partial reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1957, 70, 451–452.

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, 51, 367–371.

D'AMATO, M. R., & JAGODA, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, 64, 117–122.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, Rinehart, & Winston, 1960.

GRANDAGE, A. Orthogonal coefficients for unequal intervals. *Biometrics*, 1958, 14, 287–289.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141–154.

HARRIS, P., & NYGAARD, J. E. Resistance to extinction and number of reinforcements. *Psychol. Rep.*, 1961, 8, 233–234.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, 61, 111–121. (a)

JENKINS, H. M. Resistance to extinction when partial is followed by regular reinforcement. Paper read at Psychonomic Society, New York, 1961. (b)

KING, R. A., WOOD, P., & BUTCHER, J. *Decreased* resistance to extinction as a function of reinforcement. *Amer. Psychologist*, 1961, 16, 468. (Abstract)

LEWIS, D. J., & DUNCAN, C. P. The effect of partial reinforcement and length of acquisition-series upon resistance to extinction of a motor and a verbal response. *Amer. J. Psychol.*, 1956, 69, 644–646.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, 55, 121–128.

MARGULIES, S. Response duration in operant level, regular reinforcement, and extinction. *J. exp. Anal. Behav.*, 1961, 4, 317–321.

MURILLO, N. R., & CAPALDI, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, 61, 345–349.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, 6, 227–234.

SENKO, M. G., CHAMP, R. A., & CAPALDI, E. J. Supplementary report: Resistance to extinction of a verbal response as a function of the number of acquisition trials. *J. exp. Psychol.*, 1961, 61, 350.

THEIOS, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, 64, 1–6.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WILSON, J. J. Level of training and goal-box movements as parameters of the intermittent reinforcement effect. Unpublished doctoral dissertation, New York University, 1958.

# T MAZE REVERSAL LEARNING AFTER SEVERAL DIFFERENT OVERTRAINING PROCEDURES [1]

WINFRED F. HILL, NORMAN E. SPEAR

*Northwestern University*

AND KEITH N. CLAYTON

*Vanderbilt University*

Several investigators (Brookshire, Warren, & Ball, 1961; Capaldi & Stevenson, 1957; North & Clayton, 1959; Pubols, 1956; Reid, 1953) have demonstrated that overlearning of a discrimination facilitates its subsequent reversal. This effect at first seems paradoxical, since it appears to imply that increasing the number of reinforced trials to one cue weakens the tendency to respond to that cue.

Several mechanisms have been suggested which might explain this overlearning-reversal effect (ORE). (a) The additional practice in making choices during the overtraining may facilitate subsequent reversals. This might be mediated by acquired observing responses, as suggested by Reid (1953) and Pubols (1956). (b) The long series of rewards may make a change more discriminable and hence make reversal learning faster, a suggestion made by Capaldi and Stevenson (1957). (c) The greater number of rewards, by building up stronger $r_G$'s, may result in greater frustration and hence greater disruption when reward ceases to follow the accustomed response. This is the explanation offered by North and Stimmel (1960) for their finding that overlearning facilitated extinction in a straight alley. It is also consistent with Birch, Ison, and

Sperling's (1960) finding that with single stimulus presentation the overlearning-reversal effect is mainly attributable to the rate of extinction of the old response. (d) The long series of trials on which nearly every response is rewarded may reduce, through any of various mechanisms, the tendency to avoid the incorrect cues, thus making it easier to approach those cues when they become correct. This explanation is suggested by D'Amato and Jagoda (1961). (e) So much stimulus satiation may be built up to the correct cues that there is a tendency to avoid these cues as soon as they cease to be associated with reward. Though this has not been suggested as an explanation for the ORE, it seems consistent with research summarized by Glanzer (1958).

The present study attempted to provide evidence concerning the relative importance of these various mechanisms for the ORE. Four groups were compared. Two were the groups found in any study of the ORE: Group N reversed as soon as the discrimination was learned, Group Fr given overtraining. The other two groups were also given overtraining, but with special features. Group Co was forced to the correct side on all its overtraining trials, thus being deprived of the opportunity for making choices. Group In was given twice as many overtraining trials as Groups Fr and Co, with half forced to the correct and half to the in-

correct side. This group was thus given more experience with the incorrect side than any of the others. It was predicted that Group Fr would reverse fastest and either Group N or In slowest, with the exact positions of the groups throwing light on the various explanations of the ORE.

## EXPERIMENT I

### Method

*Subjects and apparatus.*—The *S*s were 64 experimentally naive female albino rats of the Sprague-Dawley strain, 74 to 75 days old at the beginning of experimental training. One *S* was discarded because of apparatus failure and was replaced by another *S* of the same description. The apparatus was the narrow T maze described by Cotton, Lewis, and Metzger (1958). This is an enclosed T maze with a 4-ft. stem and 9-in. arms (excluding goal box). The wooden doors beyond the choice-point, utilized to prevent retracing, were also used for forcing purposes in this experiment.

*Prehandling.*—Each *S* received six daily 3-min. sessions of prehandling, the last one 48 hr. prior to the beginning of experimental training. During each session *S* was allowed to explore a large unpainted wooden box, presented with four of the pellets later to serve as reward, and picked up and replaced at least five times by *E*. Pellets uneaten were returned with *S* to the home cage along with a fifth pellet. A once-daily feeding schedule began on the first day of prehandling and was maintained throughout experimental training. The ration was 10 gm. of finely ground Purina lab chow and was presented 30–35 min. after the start of prehandling or experimental training.

*Experimental training.*—On each trial, after placement into the start box, the door was opened as soon as *S* was oriented toward it. Located just outside the start box was a treadle which, when *S* stepped on it, started two 1/100 sec. clocks. The first clock stopped automatically when *S* stepped on a treadle located just before the choice-point. The second clock stopped when *S* stepped on a treadle just beyond the choice-point in either arm of the maze. The side designated as correct for a particular *S* was baited with two 45-mg. Noyes pellets presented in a slightly bent tin lid. No lid was present on the nonrewarded side. For half the *S*s, the right side was designated correct; for the

others the left was correct. The noncorrection procedure was used throughout, and the duration of goal-box confinement was approximately 10 sec. regardless of choice. Between trials *S* was kept for 15 sec. in a carrying cage before being placed in the start box for the next trial.

*General design.*—There were three stages to the experimental training—acquisition, overtraining, and reversal. The experimental variable was manipulated only in the overtraining stage. However, prior to the beginning of acquisition, *S*s were randomly assigned to four experimental groups of 16 *S*s each. The three stages of training required a total of 13 days for all groups.

*Acquisition.*—Twelve free trials were given on each of 3 consecutive days. Groups Fr, Co, and In received acquisition on Days 1, 2, and 3 of acquisition. Group N, however, received acquisition on Days, 1, 2, and 11.

*Overtraining.*—On Days 3–6 and 8–10 Group N was treated exactly as it had been during prehandling and was given no T maze experience during this time. It thus served as a no-overtraining control group. Group Fr was given 15 free trials on each of Days 4–6 and 8–10 and 12 free trials on Day 11, making a total of 102 free overtraining trials. Group Co was given 15 trials, all forced to the correct side, on each of Days 4–6 and 8–10, and 12 trials, also forced correct, on Day 11. This made a total of 102 forced-correct overtraining trials. Group In was given 30 trials on each of Days 4–6 and 8–10 and 24 trials on Day 11. All of these 204 trials were forced, half to the correct side and half to the incorrect. The distribution of correct and incorrect forced trials was determined randomly, and was different for each *S*. On Day 7, all *S*s in all groups were fed on schedule but were not handled.

*Reversal.*—All *S*s received 15 free trials on each of Days 12 and 13. During this stage, reward was placed in the goal box opposite to that which had originally held the reward during acquisition and overtraining, so that the formerly correct goal box became incorrect and vice versa.

### Results

*Acquisition.*—The mean proportion of correct choices on the last five trials was .93. An analysis of variance on the latter half of acquisition revealed no significant differences among the groups ($F = 1.87$). In Group N there was no evidence of forgetting

TABLE 1

TOTAL CORRECT CHOICES
IN REVERSAL

| Group | Experiment I | | Experiment II | | Experiment III | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| N | 21.88 | 2.92 | 20.31 | 5.69 | 18.15 | 7.40 |
| Fr | 18.94 | 1.48 | 16.69 | 5.42 | 9.07 | 9.13 |
| Co | 21.31 | 2.75 | 21.38 | 4.50 | | |
| In | 10.75 | 5.23 | 7.06 | 6.10 | | |

over the interval between the second and third days of training.

*Overtraining.*—Group Fr made a total of 3.3% incorrect responses during overtraining. The median number of correct responses after the last error was 54.

*Reversal.*—The course of reversal learning for the four groups is shown in Fig. 1. The mean number of correct choices on the 30 reversal trials is shown for the four groups in Table 1. The overall $F$ for the four groups was 37.01, significant at the .001 level. Adjacent groups were compared with $t$ tests. Groups N and Co did not differ significantly ($t = .57$). However, Group Co was superior to Group Fr at the .01 level ($t = 2.96$), and Group Fr in turn was superior to Group In at the .001 level ($t = 6.02$).

The mean number of incorrect responses before the first correct reversal response is shown for the four groups in Table 2. All differences were significant at the .01 level except that between Groups N and Co, which did not approach significance.

In view of the failure to find an overlearning-reversal effect in terms of total correct responses in reversal, the reversal data were also analyzed in terms of a criterion of 18 correct choices in 20 successive trials, with not more than one error in the last 10 of the 20. This is the same as Pubol's (1956) criterion except that his 20-trial units always involved 2 complete days of running, whereas ours could begin on any trial. Comparison with Pubol's second experiment is most appropriate since it was closest to ours in the kind of discrimination involved. The numbers of $Ss$ reaching this criterion in the four groups were 10, 7, 10, and 1, respectively. These values yield a $\chi^2$ of 13.7, significant at the .01 level for 3 $df$.

*Speeds.*—Speeds in feet per second on the first five trials of reversal are shown in Table 3. Stem speeds are from the starting treadle to the treadle just before the choice-point; total speeds are from the starting treadle to one of the two treadles just beyond the choice-point. For stem speeds there were no significant differences among the groups. The overall $F$ was 1.29, and the $t$ ratio between the two extreme groups (1 and 2) was 1.79. For total speeds, however, the overall $F$ was 4.16, significant at the

TABLE 2

NUMBER OF ERRORS BEFORE FIRST CORRECT
RESPONSE IN REVERSAL

| Group | Experiment I | | Experiment II | | Experiment III | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| N | 3.19 | 1.83 | 6.44 | 5.75 | 5.08 | 3.35 |
| Fr | 6.19 | 1.17 | 9.31 | 5.93 | 11.93 | 8.50 |
| Co | 3.81 | 3.02 | 5.25 | 3.77 | | |
| In | 12.44 | 5.45 | 16.19 | 8.90 | | |

TABLE 3

MEAN SPEEDS IN FEET PER SECOND ON FIRST
FIVE REVERSAL TRIALS

| Group | Experiment I | | Experiment II | | Experiment III | |
|---|---|---|---|---|---|---|
| | Stem | Total | Stem | Total | Stem | Total |
| N | 4.01 | 2.70 | 3.87 | 2.89 | 3.91 | 2.86 |
| Fr | 4.42 | 3.50 | 4.36 | 3.53 | 4.20 | 3.23 |
| Co | 4.11 | 3.03 | 4.45 | 3.48 | | |
| In | 4.07 | 3.10 | 3.44 | 2.78 | | |

.01 level. Of the differences in total speed among adjacent groups, only that between Groups Fr and In reached the .05 level of significance by *t* test.

## Experiment II

The failure to find an ORE in Exp. I brought the purpose of the study into question. An attempt was therefore made to change the conditions so as to replicate the earlier findings of an ORE. In order to bring Exp. II closer to these earlier studies, the number of overlearning trials was increased, the intertrial interval was made longer, and various other details were changed.

### Method

*Subjects and apparatus.*—The *S*s were 64 experimentally naive female albino rats of the Sprague-Dawley strain, 74 to 78 days old at the beginning of experimental training, divided into four groups of 16 as in Exp. I. Two *S*s were discarded because of apparatus failure and 5 were discarded because on three consecutive trials they failed to trip the first treadle within the 2-min. time limit or failed to trip one of the second treadles within the 3-min. limit. (On such trials, *S* was placed in a randomly selected goal box on the first occasion and subsequent odd-numbered occasions, and in the alternative goal box on all even-numbered occasions. The time score was recorded as 180 sec. and the goal box into which *S* was placed was recorded as chosen.) These discarded *S*s were replaced by *S*s of the same description. The apparatus was the same as that used in Exp. I.

*Prehandling.*—Prehandling was the same as in Exp. I, but the feeding schedule was changed slightly. The once-daily ration of 10 gm. was presented 2 hr. after the start of prehandling or experimental training, since the altered procedure increased the total time of experimental training each day from 15–25 min. to 90–110 min.

*Experimental training.*—Except for a change in intertrial interval and magnitude of reward, all details were the same as in Exp. I. In Exp. II, 3–4 min. elapsed between the placement of *S* in its carrying cage following a trial and its removal at the beginning of the subsequent trial. In addition, only one 45-

mg. pellet was present in the proper goal box on each trial. These changes applied to acquisition, overtraining, and reversal.

*Acquisition.*—Fifteen free trials were given on each of 2 consecutive days. Groups Fr, Co, and In all received acquisition on Days 1 and 2 of experimental training. Group N was prehandled (in the same manner as before) on Days 1–9 and then given acquisition on Days 10–11.

*Overtraining.*—Group Fr was given 20 free trials on each of Days 3–6 and 8–9 and 15 free trials on each of Days 10–11, for a total of 150 free overtraining trials. Group Co was given 20 trials, all forced to the correct side, on each of Days 3–6 and 8–9 and 15 trials, also forced correct, on Days 10–11, for a total of 150 forced correct trials. Group In was given 40 trials on each of Days 3–6 and 8–9 and 30 trials on Days 10–11. Half of these 300 trials were forced correct and half forced incorrect.

*Reversal.*—On Days 12 and 13 all *S*s were given reversal training as in Exp. I.

### Results

*Acquisition.*—The mean proportion of correct choices on the last five trials of acquisition was .93. An analysis of variance of the last half of acquisition revealed no significant difference among the groups ($F < 1$). To check comparability with one previous study using spatial discrimination, it was determined for each *S* on what trial, if at all, it reached the criterion defined in Exp. I. For the four groups combined, 35 of the 64 *S*s reached this criterion, the median for completion of the criterion being Trial 29. This means that by Pubols' definition the median *S* received one trial of overtraining during acquisition.

*Overtraining.*—Group Fr made a a total of 2.7% incorrect choices during overtraining. The median number of correct responses after the last error was 100.

*Reversal.*—The course of reversal learning for the four groups is shown in Fig. 1, and the mean number of
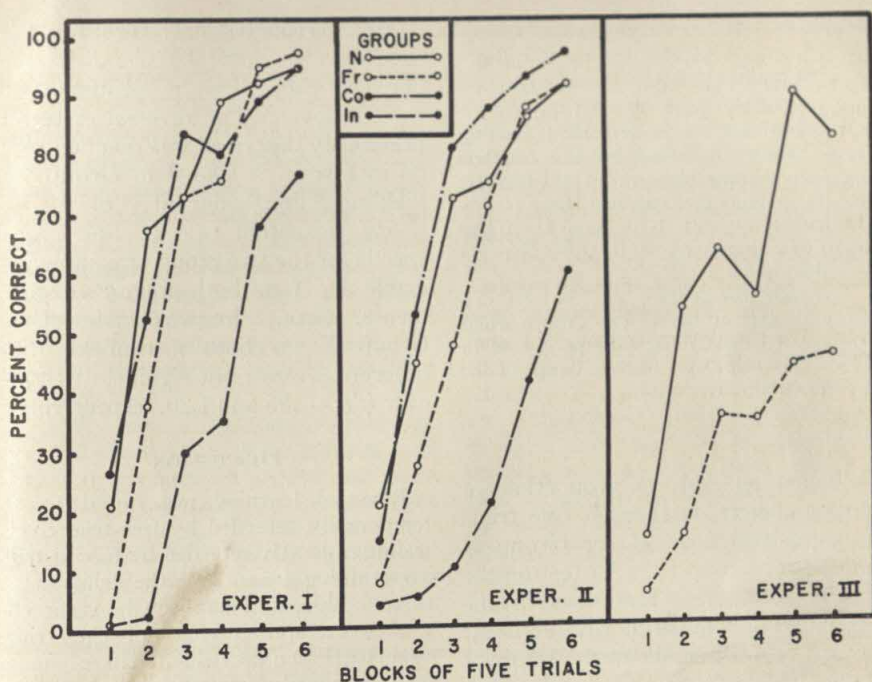
FIG. 1. Reversal learning curves in the three experiments.

correct choices for all 30 trials is shown in Table 1. The pattern of group differences is clearly similar to that found in Exp. I. Analysis of variance of the four group means yielded an $F$ of 22.77, significant at the .001 level. By $t$ test, Group In was found to be lower than each of the others at the .001 level, while Group Fr was lower than Group Co at the .05 level. Other differences were not significant.

The mean number of errors before the first reversal is shown in Table 2. Group N does not differ significantly from Groups Fr and Co, but all other differences are significant at least at the .05 level.

The reversal data were also analyzed according to the same criterion as in Exp. I. The numbers of $S$s reaching criterion in the four groups were 8, 7, 14, and 1, respectively,

yielding a $\chi^2$ of 21.3, significant at the .001 level for 3 $df$.

*Speeds.*—Speeds on the first five reversal trials are shown in Table 3. The overall difference among the groups is significant at the .01 level both for stem speeds ($F = 7.27$) and for total speeds ($F = 5.95$). In both cases the two higher groups differ significantly from the two lower, but Group Fr does not differ significantly from Co or Group N from In.

EXPERIMENT III

Since the pattern of results in Exp. II did not differ appreciably from that in Exp. I, a further attempt was made to find an ORE. Since $S$s in Exp. I and II were younger than those in some of the previous studies that found an ORE, older $S$s were used in Exp. III. Only Groups N and Fr were included in this experiment.

## Method

*Subjects and apparatus.*—The *S*s were 27 experimentally naive female albino rats of the Sprague-Dawley strain, 120–121 days of age at the beginning of experimental training. Three *S*s were discarded because on three consecutive trials they failed to trip the treadles within the specified time limits, which were the same as in Exp. II. These discarded *S*s were replaced by other *S*s of the same description.

*Procedure and design.*—The *S*s were randomly assigned to one of two groups which were treated exactly like Groups N and Fr of Exp. II. There were 13 *S*s in Group N and 14 *S*s in Group Fr.

## Results

*Acquisition.*—Mean proportion of correct choices on the last five trials of acquisition was .94 for Group N and .84 for Group Fr. A *t* test on the number of correct responses in the latter half of acquisition revealed that Group N was significantly superior at the .02 level ($t = 2.63$). No reason for this difference is apparent. The median *S* in the two groups combined just reached criterion, with no additional trials.

*Overtraining.*—Group Fr had a mean of 2.0% incorrect choices during overtraining. The median number of correct responses after the last error was 102.

*Reversal.*—The course of reversal learning is shown in Fig. 1. The total numbers of correct responses in reversal are given in Table 1. The superiority of Group N is significant at the .01 level ($t = 2.89$). In view of the superiority of Group N in acquisition, this might be attributed to a chance superiority in learning ability of *S*s in the group. However, the within-groups correlation between number of correct choices in the latter half of acquisition and number correct during reversal is negative and nonsignificant ($r = -.22$), which makes such an interpretation implausible.

Mean errors before the first reversal are shown in Table 2. The difference between the groups is significant at the .01 level. The reversal criterion previously described was reached by 6 *S*s in Group N and 3 in Group Fr, yielding a nonsignificant $\chi^2$ of .91.

*Speeds.*—Stem ·speeds and total speeds for the two groups are shown in Table 3. As in both previous experiments, Group Fr was faster than Group N on both measures. The differences were not significant, however (*t*'s = .85 and 1.20, respectively).

## DISCUSSION

Reversal learning was found to be temporarily retarded by free-trial overtraining, greatly retarded by forced-trial overtraining when half the trials were to each side, and unaffected by forced-trial overtraining when all the trials were to the correct side. These effects were primarily accounted for by the duration of perseveration on the formerly correct side in the different groups at the beginning of reversal, as may be seen by comparing Tables 2 and 3.

The most noteworthy aspect of these findings is the failure to confirm the finding of other investigators that overlearning facilitates reversal. We seem to have ruled out the possibility that the discrepancy between our results and those of previous *E*s is due to age of *S*s, number of overlearning trials, or intertrial interval. Four differences remain between our procedure and earlier ones.

1. We used a constant number of acquisition trials before overtraining instead of carrying all *S*s to a criterion. Because of the variability of the *S*s, a training procedure which gives a constant number of trials may fail to insure that all *S*s will master the original task or may permit a few to receive a degree of overtraining. Either of these conditions might facilitate reversal performance of Group 1 and reduce the chance of demonstrating an ORE. For this reason comparisons were made among three subgroups of Groups N and Fr:

(*a*) those that failed to reach the acquisition criterion, (*b*) those that had 5 or fewer acquisition trials after reaching criterion, and (*c*) those that had 6 to 10 (the maximum possible) acquisition trials beyond criterion. In all three subgroups in all three experiments, Group N made more correct reversal responses than Group Fr. (In view of the small $N$ in these subgroups, no statistical tests were made.) This suggests that the failure to use an acquisition criterion was not critical. There might be an interaction, however, between this consideration and the next one.

2. We gave a constant number of reversal trials rather than carrying each $S$ to a reversal criterion. Perhaps $Ss$ in Group Fr, once they overcame their initial disadvantage, would have surpassed Group N in reaching criterion. Neither the reversal learning curves nor the number of $Ss$ reaching criterion in the two groups lend any support to this view. As for the possible interaction of Differences 1 and 2, of the $Ss$ in all three experiments that reached the acquisition criterion with five or fewer trials to spare, 5 in Group N and 4 in Group Fr reached the reversal criterion.

3. We tried to control differences in amount of handling between overtrained and nonovertrained groups by giving Group N extra handling when they were not being run. This may have controlled away the whole effect. If this should be the explanation, it would make the overlearning-reversal effect a more trivial phenomenon than has been suspected.

4. Our task was a simple spatial discrimination with no irrelevant visual cues (such as Pubols had) and no separation of place and response cues (as with Brookshire et al.). If observing responses are the crucial factor in the overlearning-reversal effect, such a simple discrimination may be too easy for such responses to be important. This explanation implies that our task is easier than those in which the effect has been found. In terms of acquisition rate, this appears to be true of most of the other studies, though not of Pubols' Exp. 2.

If this exception can be explained by the extensive pretraining in Pubols' experiment leading to faster learning, then the simplicity of our discrimination might well be the crucial factor in our failure to find an overlearning-reversal effect. This would be consistent with D'Amato and Jagoda's (1962) failure to find an overlearning-reversal effect in a spatial discrimination. Their study, of which we were unaware when conducting ours, is to our knowledge the first published failure to find an overlearning-reversal effect with high overlearning. The situation is complicated, however, by an unpublished study of Erlebacher (1961). He failed to find the overlearning-reversal effect, even though he used a brightness discrimination. The necessary conditions for obtaining an overlearning-reversal effect thus remain in doubt.

Our results and the strikingly similar findings of D'Amato and Jagoda seem to be clearly incompatible with the first three interpretations of discrimination reversal listed in the introduction. These findings can, however, be reconciled with Interpretations 4 and 5, both of which are concerned with avoidance of the formerly incorrect side. Either of these interpretations is consistent with the rank order of our three overtraining groups. The unexpected superiority of Group N in reversal could then be explained by their weaker tendency to approach the formerly correct side, a tendency supported by the generally slower speeds of Group N. The similar patterns of stem speed and total speed suggest that time spent in the stem rather than time spent in the choice area or arms was the major source of variance in both speed measures.

## SUMMARY

Experiment I attempted to find the cause of the overlearning-reversal effect by comparing T maze reversal learning by four groups of rats that received different patterns of overtraining in acquisition. Reversal was fastest for the group receiving no overtraining and the group receiving all its overtraining trials forced to the correct side. Free-choice

overtraining gave somewhat slower reversal, and overtraining with an equal number of forced trials to the two sides gave much slower reversal.

In view of the failure to replicate previous findings of faster reversal after overtraining, two further experiments were run in an attempt to replicate these earlier findings. Both experiments gave the same pattern of results as Exp. I; no overlearning-reversal effect was found. These results appear to be consistent with interpretations in terms of stimulus satiation or of avoidance of non-rewarded cues, but not with interpretations in terms of observing responses, discriminability, or frustration.

## References

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, 60, 36–40.

Brookshire, K. H., Warren, J. M., & Ball, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, 54, 98–102.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, 50, 195–198.

Cotton, J. W., Lewis, D. J., & Metzger, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, 51, 336–341.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *J. exp. Psychol.*, 1961, 61, 45–50.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, 64, 117–122.

Erlebacher, A. Reversal learning in rats as a function of percentage reinforcement and degree of overlearning. Unpublished doctoral dissertation, University of Wisconsin, 1961.

Glanzer, M. Curiosity, exploratory drive, and stimulus satiation. *Psychol. Bull.*, 1958, 55, 302–315.

North, A. J., & Clayton, K. N. Irrelevant stimuli and degree of learning in discrimination learning and reversal. *Psychol. Rep.*, 1959, 5, 405–408.

North, A. J., & Stimmel, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, 6, 227–234.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, 49, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, 46, 107–112.

# MONETARY INCENTIVE AND RANGE OF PAYOFFS AS DETERMINERS OF RISK TAKING [1]

## LEONARD KATZ

*University of Massachusetts*

Myers and associates (Myers & Fort, 1961; Myers & Katz, 1962; Myers & Sadler, 1960) have been concerned with the effects of parameters of the payoff distribution upon the choice between gambling and not gambling. Myers and Sadler (1960) varied the number of chips which might be won or lost on each gamble (range), the average payoff being zero. When the alternative to gambling was the sure gain of one chip, gambling increased with increases in range; when the alternative to gambling was the sure loss of one chip, gambling decreased with an increase in range.

In the present experiment chips worth no money and chips worth 5¢ were used to provide data on effects of incentive value. Monetary incentive and range of payoffs are similar to each other in that an increase in either increases the risk associated with each gamble. The implied hypothesis is that increased incentive, like increased range, may result in more gambling, where the alternative to gambling is the sure gain of one chip, and in less gambling when the alternative to gambling is the sure loss of one chip. The objective of the present study was, therefore, to obtain data to test this hypothesis. The S's decision to gamble under

different payoff ranges was followed by the loss or gain of chips which had no monetary value or was followed by the loss or gain of chips worth 5¢.

In addition to providing data on the effects of monetary incentive, the present study investigated the effects of a payoff range greater than that used in the Myers and Sadler (1960) study. To control for any range effects due to differences among ranges in sequences of payoffs, one sequence was randomly generated and the other two were derived from it in a manner described in the procedure section.

## METHOD

*Materials.*—Four decks of 100 3 × 4 in. white cards were prepared. The known payoff deck contained 50 cards with +1 written on them, alternated randomly with 50 cards with −1. The other decks, of 100 cards each, provided for three different ranges of unknown payoffs. The narrow-range deck (N), had integers randomly chosen from +2 to +6 and from −2 to −6. The medium range deck (M), was constructed by adding 10 to every positive number of (N) and subtracting 10 from every negative number, giving a range of +12 to +16 and −12 to −16, retaining the same ordinal positioning of cards as found in Deck N. The wide range deck (W), was constructed by adding and subtracting 20 to the integers of Deck N.

*Procedure.*—On each of 3 successive days S was presented with the known payoff deck, and a different one of the three unknown payoff decks. Order of use of Decks N, M, and W was counterbalanced in a 3 × 3 Latin square with 6 Ss given each order. Eighteen Ss gambled only for poker chips (0¢) and 18 others gambled for chips worth 5¢ apiece. The main features of the experimental design are shown on the left-hand side of Table 1.

The Ss were tested individually, being given full instructions at the beginning of Session 1. For Sessions 2 and 3 Ss were told

## TABLE 1

MEANS OF PROPORTIONS OF GAMBLING RESPONSES ON $+1$ AND $-1$ TRIALS
FOR SUCCESSIVE 25-TRIAL BLOCKS

| Incentive | Alternative to Gambling | Range | Trials | | | | Overall Trials |
|---|---|---|---|---|---|---|---|
| | | | 1–25 | 26–50 | 51–75 | 76–100 | |
| 0¢ | +1 | N | .594 | .539 | .456 | .511 | .525 |
| | | M | .620 | .551 | .494 | .604 | .567 |
| | | W | .466 | .472 | .394 | .596 | .482 |
| | −1 | N | .903 | .910 | .874 | .911 | .899 |
| | | M | .796 | .796 | .822 | .744 | .798 |
| | | W | .838 | .836 | .756 | .800 | .797 |
| 5¢ | +1 | N | .654 | .588 | .489 | .563 | .573 |
| | | M | .688 | .537 | .461 | .574 | .565 |
| | | W | .658 | .574 | .483 | .689 | .601 |
| | −1 | N | .833 | .893 | .881 | .878 | .871 |
| | | M | .796 | .790 | .815 | .761 | .790 |
| | | W | .782 | .756 | .741 | .750 | .757 |

only that the unknown payoff deck of that day was a new one. Two hundred poker chips were stacked in front of $S$, and $S$s with a monetary incentive (5¢) were told that each chip was worth 5¢; i.e., their initial stake was worth $10.00.

Briefly, $S$s were instructed to turn over the top card in the deck of known payoffs at the beginning of each trial. They then chose between standing pat by accepting the gain or loss of one chip represented by the card and gambling by drawing the top card from the deck of unknown payoffs. If $S$ decided not to gamble, the top card in the unknown payoff deck was turned anyway, showing what he would have won or lost had he gambled.

*Ratings.*—At the end of each session, $S$s were given an 11-point scale from $-5$ to $+5$, along which they rated the means of the known payoff deck and of the unknown payoff deck used that session. They were also asked to describe their gambling strategies and changes in strategy. After Session 3, $S$s were asked what they would have done differently in Sessions 1 and 2.

*Subjects.*—The $S$s were 36 male undergraduates enrolled in the university summer session, who were divided randomly into two groups of 18 $S$s each. Each $S$ was paid $3.00.

## RESULTS

*Choices.*—The scores were proportions of choices to gamble on both $+1$

and $-1$ trials during each of four 25-trial blocks. For example, if in a particular block $S$ gambled on 3 of the 12 (or 13) trials in which the alternative to gambling was $+1$, his $+1$ proportion for that block was .250 (or .231). Table 1 presents means of these proportions for each combination and range. An analysis of variance was performed on arc-sine transforms of the proportions. The results of this analysis (shown in Table 2), in conjunction with the relationships shown in Table 1, suggest the following conclusions: (a) under all combinations of range and incentive, more risk-taking occurs on $-1$ trials than on $+1$ trials ($P < .001$); (b) the difference in incentives had little effect on total number of risks taken; (c) more gambling occurred with Deck M than with Deck N on $+1$ trials and the reverse on $-1$ trials ($P < .01$); and (d) value, incentive, and range have a joint effect on gambling ($P < .001$). A further breakdown suggests that this effect is due largely to differences in the quadratic curvature of the V-I

curves over the three ranges. For +1 trials, the curves for incentive are essentially mirror images of each other; for −1 trials, the curves for incentive are essentially parallel to each other.

In addition, there were a number of significant interactions involving blocks of trials. These effects are probably due to the interaction of range, incentive, and value with the effects of both the temporal sequence of blocks within a single session and the different mean payoffs of blocks.

*Ratings.*—Table 3 shows the means of *S*s' ratings of the means of the known and unknown payoff decks for each combination and incentive. The means for the known payoff deck were closer to the true mean of zero and

### TABLE 3
#### MEANS OF *S*s' RATINGS OF DECK MEANS

| Range | Payoff | | | |
|---|---|---|---|---|
| | Known | | Unknown | |
| | 0¢ | 5¢ | 0¢ | 5¢ |
| N | .33 | .16 | .44 | −0.66 |
| M | .33 | .16 | .00 | 1.05 |
| W | .44 | −.22 | .88 | 0.88 |

were less variable than the means for the unknown payoff decks. Neither set of means varied systematically with incentive or range of unknown payoff. Nor were there any systematic relationships between these deck means and either overall proportions of risks or proportions of risks for the last 25 trials. Most *S*s reported following a "gambler's fallacy"[2] strategy and none reported an awareness of gambling differentially with different decks of unknown payoffs. Incentive had no differential effect.

### TABLE 2
#### ANALYSIS OF VARIANCE OF ARC-SINE TRANSFORMS OF THE PROPORTION OF GAMBLING RESPONSES IN EACH TRIAL

| Source | df | MS | F |
|---|---|---|---|
| Incentive (I) | 1 | 710.00 | |
| *S*s/I | 34 | 2,047.71 | |
| Value (V) (known payoff) | 1 | 103,825.00 | 33.72*** |
| Range (R) | 2 | 1,361.40 | 2.22 |
| Blocks (B) | 3 | 1,454.93 | 4.90** |
| V × B | 3 | 987.23 | 3.28* |
| V × R | 2 | 1,835.60 | 8.73** |
| R × B | 6 | 187.01 | 8.87** |
| R × B × V | 6 | 367.75 | 2.25* |
| V × I | 1 | 2,910.50 | 1.07 |
| B × I | 3 | 56.80 | 0.19 |
| R × I | 2 | 292.70 | 0.47 |
| V × I × R | 2 | 1,606.10 | 13.66*** |
| R × I × B | 6 | 61.66 | 2.82* |
| V × I × B | 3 | 238.16 | 1.39 |
| R × B × V × I | 6 | 154.27 | 1.57 |
| *S*s × V/I | 34 | 2,701.36 | |
| *S*s × B/I | 102 | 296.53 | |
| *S*s × R/I | 86 | 611.05 | |
| *S*s × V × B/I | 102 | 300.18 | |
| *S*s × V × R/I | 68 | 210.20 | |
| *S*s × R × B/I | 204 | 21.80 | |
| *S*s × V × B × R/I | 204 | 162.91 | |

\* $P < .05$.
\*\* $P < .01$.
\*\*\* $P < .001$.

## DISCUSSION

The decision to gamble or not to gamble following known outcomes of loss or gain of a poker chip was investigated as a function of three ranges of unknown payoff involving the loss or gain of chips worth nothing or worth 5¢. Previous findings (Myers & Katz, 1962; Myers & Sadler, 1960) of more gambling when the alternative to gambling was the loss of a chip than when the alternative was the gain of a chip were confirmed. This difference was reduced by monetary incentive, although the interaction was not statistically significant: trials when the alternative was −1 were followed by fewer choices to gamble for chips worth 5¢ than for those worth nothing, while trials where the alternative was +1 were followed by more choices to gamble for chips worth 5¢ than for those

[2] Due to the procedure for constructing the payoff sequences, prediction of the alternative event following a run is not really fallacious.

worth nothing. Thus, there is some suggestion that monetary payoff and range are functionally equivalent; for all three ranges monetary incentive yielded an increase in gambling on $+1$ trials, a decrease on $-1$ trials.

Other experiments have shown the form of value and range interaction obtained for the 5¢ incentive group. Myers and Sadler, using three ranges up to and including the range of Deck M of the present study, found that gambling consistently increased as range increased when the alternative was a gain of one chip, but gambling decreased as range increased when the alternative was a one-chip loss. Myers and Katz (1962) obtained similar results through a range of Deck M. Suydam and Myers (1962), using a very different procedure, found this convergence of positive and negative value curves over range, for several values. It may be assumed that the gamble represents an approach-avoidance conflict, both tendencies increasing as range does. Within this frame of reference, the data suggest that, as range increases, the avoidance gradient rises more swiftly against negative alternatives; the approach gradient rises more swiftly against positive alternatives.

## SUMMARY

The effects on gambling behavior of monetary incentive, range of payoffs for gambling, and value of a payoff which could be taken in lieu of gambling were determined. One group of 18 Ss gambled for chips only and another group of 18 Ss gambled for chips worth 5¢ each. Three ranges of unknown payoffs were used, one at each of three sessions. The known payoff, the acceptance of which was the alternative to gambling, remained constant.

Neither incentive nor range had a significant effect upon the total number of risks taken. The Ss gambled significantly more when the alternative to gambling was a loss ($-1$) than when it was a gain ($+1$). Several interactions were significant, including Range × Value and Range × Value × Incentive. These led to the conclusion that gambling is affected differentially on $+1$ and $-1$ trials by the range of chips to be gained or lost, and by the interaction of range and monetary incentive.

## REFERENCES

MYERS, J. L., & FORT, G. G. A sequential analysis of gambling behavior. Paper presented at Psychonomic Society, Columbia University, 1961.

MYERS, J. L., & KATZ, L. Range of payoffs and feedback in risk-taking. *Psychol. Rep.*, 1962, 10, 483–486.

MYERS, J. L., & SADLER, E. Effects of range of payoffs as a variable in risk taking. *J. exp. Psychol.*, 1960, 60, 306–309.

SUYDAM, M. M., & MYERS, J. L. Two parameters of risk-taking behavior. *Psychol. Rep.*, 1962, 10, 559–562.

# SPATIAL S-R CONTIGUITY IN HUMAN DISCRIMINATION LEARNING

## C. D. STANDISH AND R. A. CHAMPION

*University of Sydney*

Though differing in purpose, both of the experiments to be reported are basically similar in design to an earlier study (Champion & Standish, 1960) in which preliminary training was given with two pairs of stimuli in fixed spatial relations (spatial discrimination learning) followed by a test stage involving the two pairs and their transposes (nonspatial learning). Under these conditions negative transfer from training to test occurred when between-pair differences among the stimuli were greater than within-pair differences, but positive transfer occurred when within-pair differences dominated. These findings were interpreted in terms of the type of S-R theory proposed by Spence (1960); in particular, the differential transfer was explained through the presence or absence of within-pair discriminations on the part of S in the training period, it being argued that the occurrence of these discriminations promoted positive transfer while their absence caused interference. The basis of the present studies rested in the further assumption that spatial S-R contiguity would also promote within-pair discriminations, and it was predicted that groups trained and tested under such conditions would show positive transfer, even though between-pair differences were dominant.

The first experiment consisted of a repetition of the 1960 study with the exception that S was required to respond by pressing a button adjacent to the stimulus rather than one some distance from it. In order to make a severe test of the effect of this form

of S-R contiguity, the stimuli in each pair were presented close together in space (S-S proximity), but this allowed the possibility that the result of the experiment was due to proximity rather than contiguity, the two factors being confounded. The second experiment was therefore conducted in an attempt to separate these two factors and attention was concentrated on the contiguity variable, with degree of proximity held constant.

## EXPERIMENT I

### Method

*Subjects.*—The Ss were 44 undergraduates from courses in psychology at the University of Sydney, there being 13 Ss in Groups 1 and 2, and 9 in Groups 3 and 4. Color-blind students were excluded.

*Apparatus and stimuli.*—The apparatus was identical with that used previously (Champion & Standish, 1960), allowing S to be presented with colored circles of light 1 in. in diameter, except that the lights were only $\frac{1}{8}$ in. apart in the horizontal midline of the $5 \times 2$ in. milk-glass screen, and the response buttons were mounted as close to the stimuli as possible, one button being located immediately above each light. In an attempt to equate within-pair differences the stimulus settings were adjusted to the following values of red (R) and green (G) on the Munsell system, with brightness and saturation held approximately constant: Stimulus $R_1 = 5R$, $R_2 = 5YR$, $G_1 = 7.5GY$, $G_2 = 2.5GY$. The results of the 1960 study had already been confirmed with these settings (Standish, 1960).

*Procedure.*—The instructions to S were slightly modified to allow for the new location of the response buttons, and between responses S's hand rested on the table immediately below the vertical midline of the screen, but otherwise the procedure was identical with that reported earlier, the control groups being denied preliminary training

## TABLE 1
### STIMULUS PAIRS FOR Ss IN EXP. I

| Stage | Experimental Groups | | Control Groups | |
|---|---|---|---|---|
| | 1<br>b > w | 3<br>w > b | 2<br>b > w | 4<br>w > b |
| 1 | R₁+:R₂−<br>G₁−:G₁+ | G₁+:R₂−<br>G₂−:R₁+ | —<br>— | —<br>— |
| 2 | R₁+:R₂−<br>G₂−:G₁+<br>R₂−:R₁+<br>G₁+:G₂− | G₁+:R₂−<br>R₂−:G₁+<br>R₂−:G₁+<br>R₁+:G₂− | R₁+:R₂−<br>G₂−:G₁+<br>R₂−:R₁+<br>G₁+:G₂− | G₁+:R₂−<br>G₂−:R₁+<br>R₂−:G₁+<br>R₁+:G₂− |

Note.—The spatial relations of the symbols in each pair correspond to the actual left-right arrangements of the stimuli. The designations b > w etc. refer to relative magnitudes of stimulus differences between and within pairs.

so as to allow an assessment of the nature and amount of transfer. A summary of the procedure and stimulus pairs is given in Table 1.

### Results

The results, in terms of mean trials to the criterion of eight successive correct trials, are summarized in Table 2. Because of the presence of some extreme scores, nonparametric statistical tests were used. The application of a $U$ test to the data of the first stage for Groups 1 and 3 (two stimulus pairs in fixed spatial relations) confirmed the earlier finding of superior performance with between-pair differences dominant ($U = 6$ with $n_1 = 13$ and $n_2 = 9$, $P < .001$).

## TABLE 2
### TRIALS TO CRITERION IN EXP. I

| Stage | Between-Pair Differences Dominant | | Within-Pair Differences Dominant | |
|---|---|---|---|---|
| | Group 1 (Exp.) | Group 2 (Control) | Group 3 (Exp.) | Group 4 (Control) |
| Training | | | | |
| Mean | 14.4 | — | 25.9 | — |
| SD | 3.6 | — | 9.4 | — |
| Test | | | | |
| Mean | 19.0 | 32.8 | 16.3 | 33.6 |
| SD | 22.8 | 29.4 | 4.8 | 19.8 |

The use of a factorial median test (Sutcliffe, 1957) with the data of the test stage (two stimulus pairs and their transposes) failed to reveal any differential transfer, there being no significant interaction between the experimental-control variable and type of stimulus difference ($\chi^2 = 3.38$ for 1 $df$). Subsequent $U$ tests showed that significant positive transfer occurred with both between-pair and within-pair differences dominant ($U = 38.5$ with $n_1$ and $n_2 = 13$, $P < .02$, and $U = 11.5$ with $n_1$ and $n_2 = 9$, $P < .02$, respectively).

## EXPERIMENT II

### Method

Subjects.—The Ss were 60 undergraduates from courses in psychology, divided into four groups of 15. Color-blind students were not excluded.

Apparatus.—The apparatus consisted of a sheet of building board 72 in. wide and 36 in. high, mounted vertically on a table in front of S and containing two $2\frac{1}{4} \times 2\frac{1}{4}$ in. pearl-perspex squares 4 in. apart in the horizontal midline. An attempt was made to locate S equidistant from the two squares so that the centers subtended the same visual angle as did the centers of the circles in the 1960 study. Housed behind the perspex squares were 15-w. lamps whose intensity was controlled by two variacs. The duration of illumination of the stimuli on any trial was 200 msec. In order to achieve spatial contiguity of stimulus and response S was instructed to use response buttons mounted on each perspex square, one in each top corner nearer to the vertical midline of the building board. To obtain noncontiguity conditions, S was told to use two buttons mounted side by side in a small metal box located 3 in. in front of the board, the box being centered 16 in. to the right of the midline of the board. The response buttons activated indicator lamps as a signal to E.

Stimuli.—The chief basis of discrimination in Exp. II was intensity of illumination, and the four values of the stimuli in foot-Lambert units were as follows: $B_1 = 125$, $B_2 = 100$, $D_1 = 2.5$, $D_2 = 3.2$. These illumination intensities were measured with a photometer at the usual location of S's head. Lights $B_1$ and $B_2$ were similar bright stimuli, whereas $D_1$

and $D_2$ were similar dull stimuli. The change from bright to dull illumination of the filament lamps was accompanied by the usual change in hue.

*Procedure.*—The instructions to $S$ were again modified to the extent demanded by the changes in the apparatus and the conditions of training. The independent variable was degree of spatial S-R contiguity (buttons near to or far from lights) and the stimulus conditions were limited to dominance of between-pair differences, the required discriminations being between $B_1$ and $B_2$, and between $D_1$ and $D_2$. Otherwise the procedure was as before, with training and test stages for the two experimental groups (Group 1, contiguity, and Group 3, noncontiguity), and test stages alone for the corresponding control groups (Groups 2 and 4, respectively). The criterion of learning was increased to 12 successive correct trials.

## Results

The trials-to-criterion scores for the training and test stages are presented in Table 3. For these data it was possible to use a parametric statistic, and the application of a $t$ test to the training scores of Groups 1 and 3 revealed no significant difference. An analysis of the variance for the test data of the second stage showed that while the pretraining had no consistent effect (experimental vs. control groups), the influence of spatial S-R contiguity and degree of differential transfer (interaction) were both statistically significant. Follow-up $t$ tests showed that the differences

TABLE 3

TRIALS TO CRITERION IN EXP. II

| Stage | Contiguity | | Noncontiguity | |
|---|---|---|---|---|
| | Group 1 (Exp.) | Group 2 (Control) | Group 3 (Exp.) | Group 4 (Control) |
| Training | | | | |
| Mean | 25.6 | — | 29.7 | — |
| *SD* | 20.6 | — | 20.4 | — |
| Test | | | | |
| Mean | 18.9 | 45.7 | 68.5 | 57.6 |
| *SD* | 8.9 | 29.4 | 31.3 | 29.9 |

between Groups 1 and 2, and Groups 1 and 3 were significant ($t = 2.56$, $P < .02$, and $t = 4.74$, $P < .01$, respectively, for 28 $df$) but that the differences between Groups 2 and 4, and Groups 3 and 4 were not significant. It thus appears that positive transfer occurred only under contiguity conditions, but that there was no negative transfer with noncontiguity of S and R.

## DISCUSSION

The results of the two experiments lead to the conclusion that spatial S-R contiguity, like between-pair similarity, leads to positive transfer from the spatial to the nonspatial learning situation (training to test), and if the theoretical interpretation given in terms of S-R theory is correct, then this is because the contiguity promotes more effective within-pair discrimination in the first stage. The chief finding of Exp. I was that positive transfer from training to test may be obtained even when between-pair differences are dominant, provided that $S$ be required literally to approach the positive stimulus cue in the course of the response. This result is to be contrasted with that of the 1960 study, where the dominance of between-pair differences produced negative transfer. That the present outcome cannot be attributed to the greater proximity of the stimulus cues in Exp. I is demonstrated by the similar result in Exp. II with a return to the same degree of non-proximity as obtained in the 1960 study.

The data of Exp. II suggest that the direct effects of contiguity may not be as powerful with adult humans as with children and lower animals, by comparison with the marked effects obtained by Murphy and Miller (1958, 1959). Learning was more efficient under contiguity conditions both in the training of Groups 1 and 3 and in the test periods of Groups 2 and 4 (Table 3), but in neither case was the difference statistically significant. However, when the two scores for each $S$ in the experimental

groups of Exp. II (Groups 1 and 3) on the training and test stages were combined, so as to make the two tasks one, a significant difference emerged ($t = 4.32$ for 28 $df$, $P < .01$), possibly due to the greater stability of the scores. It will also be noted that the significant negative transfer obtained in the 1960 study with noncontiguity and between-pair differences dominant was not duplicated in Groups 3 and 4 of Exp. II, although the result was in the same direction.

Spence's recent elaboration of an S-R theory of selective learning (Spence, 1960) points to some of the complexities to be coped with in any detailed formulation, but brief consideration may profitably be given to the broader theoretical significance of spatial S-R contiguity. In the present context this variable has been assumed to act through within-pair discrimination, and in S-R terms the efficiency of this latter discrimination depends in turn upon the relative strengths of the orienting and approach tendencies to the positive and negative discriminanda. The necessary link between spatial contiguity and the relative strengths of the correct and incorrect responses may now be provided if it be allowed that spatial contiguity amounts to temporal contiguity of stimulus and response, for the importance of the latter variable in simple learning seems undoubted (e.g., Champion, 1962). The equation of spatial and temporal contiguity in the discrimination-learning situation follows from the fact that when $S$ is required to respond by pressing a button in or near the positive stimulus (contiguity) then there is more likely to be a short time interval between exposure to the stimulus and the occurrence of the response than if $S$ has to look away from the stimulus to locate the appropriate response button (noncontiguity). This effect might show up more clearly if the stimuli were presented for a long time interval, but the argument applies even with intervals as short as 100 msec., for

there is no requirement that stimulus and response overlap in time, the terms "contiguity" and "noncontiguity" being relative rather than absolute.

## SUMMARY

Two experiments on discrimination learning were conducted under conditions in which preliminary training was given with two pairs of stimuli in fixed spatial relations, followed by test learning involving the two pairs and their transposes. The aim was to test the effects of spatial S-R contiguity, for it was predicted that contiguity would cause positive transfer from training to test. The prediction was confirmed, and the result was interpreted in S-R terms with the hypothesis that contiguity, like between-pair similarity, promotes within-pair discriminations on the part of the learner, it being supposed that spatial contiguity has this effect through the more basic variable of temporal contiguity.

## REFERENCES

CHAMPION, R. A. Stimulus-response contiguity in classical aversive conditioning. *J. exp. Psychol.*, 1962, **64**, 35–39.

CHAMPION, R. A., & STANDISH, C. D. Stimulus differences in discrimination learning. *J. exp. Psychol.*, 1960, **60**, 78–82.

MURPHY, J. V., & MILLER, R. E. The effect of the spatial relationship between the cue, reward, and response in simple discrimination learning. *J. exp. Psychol.*, 1958, **56**, 26–31.

MURPHY, J. V., & MILLER, R. E. Spatial contiguity of cue, reward, and response in discrimination learning by children. *J. exp. Psychol.*, 1959, **58**, 485–489.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

STANDISH, C. D. Stimulus differences and response tendencies in discrimination learning. Unpublished master's thesis, University of Sydney, 1960.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, **54**, 134–137.

# SUPPLEMENTARY REPORT: STIMULUS FAMILIARIZATION IN PAIRED-ASSOCIATE LEARNING

RUDOLPH W. SCHULZ AND IRVING F. TUCKER

*State University of Iowa*

When *S*s are familiarized with the stimulus units but not the response units of a list prior to paired-associate (PA) learning, it has generally been found that stimulus familiarization (SF) has a slight inhibitory effect or no effect at all on PA performance (Underwood & Schulz, 1960). Recently, Gannon and Noble (1961) have reported significant facilitation of performance on a list of paired dissyllables following 20 trials of SF. Since the latter result is the one to be expected if frequency of prior experience is the vehicle through which stimulus meaningfulness has its effect on PA performance, it is a result with considerable theoretical significance (e.g., Cieutat, Stockwell, & Noble, 1958; Underwood & Schulz, 1960). However, the finding of a positive effect from SF is also conspicuously inconsistent with the results of previous studies. Therefore it seemed especially important to further assess the reliability of this result and consider potential alternative explanations for it.

One such alternative is that Gannon and Noble's procedure of having *S* articulate the stimulus unit during the PA anticipation interval, when combined with variation in amount of SF, may have inadvertently produced simultaneous variation in the *effective length* of the anticipation interval. Since practice in articulation of stimulus units is directly related to amount of SF, familiarized *S*s might spend relatively less of the 2-sec. anticipation interval for stimulus articulation than nonfamiliarized *S*s. On the basis of the presumed direct relationship between PA performance and length of the anticipation interval Gannon and Noble's results would then be expected. The present experiment tested this hypothesis by comparing the performance of *S*s instructed to pronounce the stimulus units during PA anticipation with performance of *S*s instructed *not* to pronounce. A significant interaction of PA instruction and amount of SF will be required to support the present contention.

*Method.*—A 2 × 3 factorial design with 2 levels of PA instruction—articulation (A) vs. nonarticulation (NA) of stimulus units—and 3 amounts of SF (0, 20, and 60 trials) was used. The six respective conditions will be referred to in terms of the values of the

independent variable associated with them (e.g., Cond. $A_0$ articulation instructions and O familiarization, Cond. $NA_{60}$ nonarticulation instructions and 60 trials of familiarization, etc.).

The materials and procedures were identical to those used by Gannon and Noble (1961) with the following exceptions: (*a*) A .85-sec. rate of presentation was used during familiarization; (*b*) PA performance consisted of 17 anticipation trials.

A total of 144 *S*s, 24 per condition, taking introductory psychology at the University of Iowa were randomly assigned to conditions as they appeared at the laboratory. The *S*s had not served in prior verbal learning experiments.

*Results and discussion.*—Performance on the PA list under the six conditions, in terms of mean total number of correct responses during 17 anticipation trials, is shown in Fig. 1. The predicted interaction between PA instructions and amount of SF was obtained, and is shown by an analysis of variance to be the only significant effect ($F = 3.78$, $df = 2/138$, $P < .05$). In agreement with the results of Gannon and Noble (1961), performance was a monotonic increasing function of number of familiarization trials when *S*s were required to pronounce the stimulus terms of a PA list prior to anticipation of the response terms. However, with
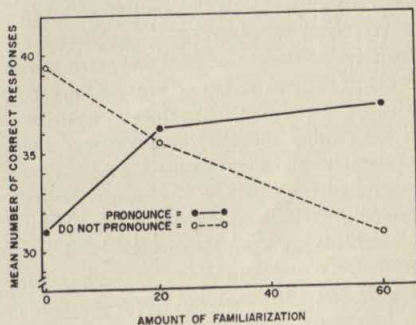


FIG. 1. Mean total number of correct responses during 17 anticipation trials as a function of PA instructions and number of stimulus familiarization trials. (The standard error of the means in Fig. 1, as estimated from the within-groups *MS* of the overall analysis of variance, was 2.71.)

nonarticulation PA instructions, performance was inversely related to amount of familiarization. It can also be seen from Fig. 1 that, even though the facilitating effects of familiarization appear to be approaching an asymptotic level under Cond. $A_{60}$, performance under Cond. $NA_0$ was slightly better than under Cond. $A_{60}$. Irrespective of how proficient $S$ becomes at pronouncing the stimulus units, it takes longer to pronounce than not to pronounce; it takes longer to say something than to say nothing.

Intercomparison of the various conditions via the critical difference technique (Linquist, 1953) revealed two significant $(P < .05)$ differences, Cond. $NA_0$ vs. Cond. $A_0$ and Cond. $NA_0$ vs. Cond. $NA_{60}$. Inspection of acquisition as a function of trials for each of the conditions did not reveal evidence of interaction between treatments and trials.

In conclusion, it is apparent that PA instructions regarding $S$'s response to the stimulus term during the anticipation interval can determine whether SF facilitates or inhibits PA performance. We believe that these effects are attributable to covariation of the effective length of the anticipation interval with amount of familiarization. Furthermore, it may be expected that such factors as word length, pronunciability, and meaningfulness will also, depending on the length of the anticipation interval, interact with PA instructions and amount of familiarization.

### REFERENCES

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness $(m, m')$ in paired-associate learning. *J. exp. Psychol.*, 1958, 56, 293–302.

GANNON, D. R., & NOBLE, C. E. Familiarization $(n)$ as a stimulus factor in paired-associate verbal learning. *J. exp. Psychol.*, 1961, 62, 14–23.

LINQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

# SUPPLEMENTARY REPORT: TIME BETWEEN PAIRINGS AND SHORT–TERM RETENTION [1]

LLOYD R. PETERSON, KENNETH HILLNER, AND DOROTHY SALTZMAN

*Indiana University*

Peterson, Saltzman, Hillner, and Land (1962) found marked forgetting within 8 sec. after a single paired-associate presentation, when other presentations filled the interval. The present study investigates the effect of an 8-sec. interval similarly filled which is inserted between the first and second presentations of an individual pair later tested for retention.

*Method.*—The technique of the previous experiment was used. The first as well as the second pairing consisted of $S$ reading aloud the stimulus and the response from the drum. Either 0 or 8 sec. separated the two pairings. The retention interval, measured from removal of the second presentation from the drum, was 2, 4, 8, or 16 sec. There were 42 $Ss$ who were tested 16 times in each of the eight conditions. Stimuli were three- and four-letter words. The responses were indicated in the instructions to be the numbers 1–10. Twelve seconds rest separated 16 blocks of 31–41 exposures, save for a 1-min. rest between Blocks 8 and 9.

*Results.*—Table 1 shows that massed pairings resulted in superior recall at the 2- and 4-sec. retention intervals, while 8-sec. spacing was superior at the 8- and 16-sec. retention intervals. An analysis of variance found significance for both Spacing $(F = 10.72, df = 1/41, P < .01)$ and Retention $(F = 71.80,$

### TABLE 1
#### PROPORTIONS CORRECTLY RECALLED

| Spacing Interval | First Half | | | | Second Half | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Retention Interval (Sec.) | | | | Retention Interval (Sec.) | | | |
| | 2 | 4 | 8 | 16 | 2 | 4 | 8 | 16 |
| 0 Sec. | .82 | .69 | .45 | .48 | .85 | .74 | .45 | .46 |
| 8 Sec. | .81 | .66 | .56 | .56 | .82 | .63 | .60 | .60 |

$df = 3/123, P < .01$). Their interaction was also significant ($F = 16.47, df = 3/123, P < .01$). The two halves of the session did not differ significantly ($F = .73, df = 1/41, P > .05$). After pooling data from the two halves of the session, individual $t$ tests showed that differences between the two spacing conditions were significant at the 4-, 8-, and 16-sec. retention intervals ($t = 3.07, 5.23, 4.84; df = 41, P < .01$).

A paradox is presented by the finding that when an interval during which marked forgetting can be shown to occur is introduced between pairings, there is improvement in retention at long intervals. Underwood's

(1961) explanation of distributed practice of lists does not seem appropriate here, since he concludes that distribution is superior to massing only when response learning is involved. Response learning was minimized by the instructions in the present study of spacing between individual pairings.

### REFERENCES

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

Underwood, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, **68**, 229–247.

(Received August 18, 1961)

# SUPPLEMENTARY REPORT: YOKED COMPARISONS OF CLASSICAL AND AVOIDANCE EYELID CONDITIONING UNDER THREE UCS INTENSITIES [1]

I. GORMEZANO, JOHN W. MOORE, AND EDWARD DEAUX

*Indiana University*

Moore and Gormezano (1961) observed that classical conditioning when experimentally equated in terms of partial reinforcement pattern and number of UCS occurrences by a yoking procedure, was inferior to avoidance conditioning. The present investigation was conducted to determine the effects of UCS intensity on such yoked comparisons of classical and avoidance conditioning.

*Method.*—The general apparatus, procedure, and stimuli were the same as in the earlier experiment. The only variation was in the intensities of the UCS employed. Twenty $Ss$ were assigned to each of the six cells of a $2 \times 3$ factorial design in which classical and avoidance conditions were made orthogonal to three UCS puffs of nitrogen which had intensities sufficient to support 40-, 80-, and 160-mm. columns of mercury. In addition 2 male $Ss$ were lost because of apparatus failure. The two recording systems and sex were also made orthogonal to the classical-avoidance and UCS intensity dimensions.

*Results.*—The distributions of response latencies for all six groups in acquisition and extinction were recorded. The distributions (not shown) revealed that aside from the

higher frequency of responses in the CR range for the avoidance groups, the only discernable difference was the tendency for the modal responses to decrease in latency, under both conditioning procedures, as UCS intensity increased.

Figure 1 presents the results of plotting percentage CRs for the six experimental groups in acquisition and extinction. The initial points on all acquisition curves are the mean percentage CRs on Trial 1 and the remaining points are for successive blocks of 10 trials. The extinction curves are plotted in 5-trial blocks. The figure indicates that the acquisition performance of the avoidance groups (A160, A80, A40), for each of the three UCS intensities, was superior to each of their respective yoked-classical groups (i.e., A160 vs. Y160, A80 vs. Y80, and A40 vs. Y40). Though Groups A160 and A80 follow essentially identical courses of acquisition, group Y160 was inferior to Y80. The extinction curves reveal that the higher level of responding of the avoidance groups in acquisition also persisted in extinction. As was observed in the previous study, CRs of the avoidance groups demonstrated steeper decay functions than the yoked-classical groups and are probably in large part due to their having started at higher levels of responding.

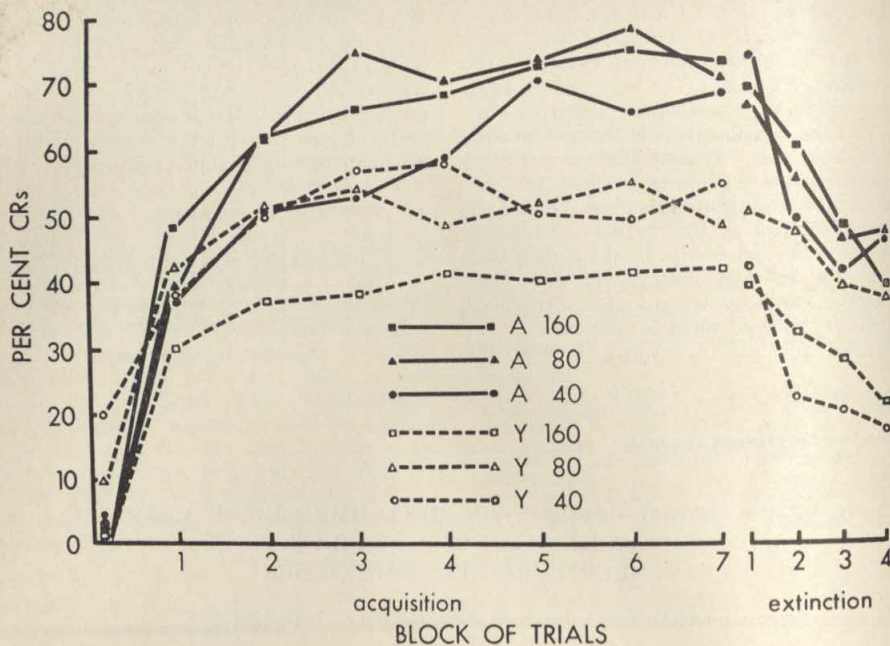Split-plot analyses of variance (Snedecor,

FIG. 1.   The percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks in extinction.

1959) on the arc-sine transform of individual percentage CRs for the 70 acquisition and 20 extinction trials revealed significant $F$ values for the classical vs. avoidance comparisons in acquisition ($F = 27.15$, $df = 1/48$, $P < .005$) and extinction ($F = 22.55$, $df = 1/48$, $P < .005$). The UCS intensity dimension failed to produce significant differences in either acquisition or extinction. A significant $F$ value was obtained for the Classical-Avoidance × UCS Intensity interaction ($F = 14.39$, $df = 2/48$, $P < .005$), reflecting the fact that as UCS intensity increased, performance of the avoidance groups increased while performance of the yoked-classical groups decreased. This interaction appears to be simply a function of a negative correlation between the performance levels of the paired $Ss$. However, the significant interaction is heavily weighted by the low level of responding of Group Y160. If the performance of Group Y160 is not sampling error, its

performance relative to Y80 suggests a Partial Reinforcement × UCS Intensity interaction. A significant $F$ value was also obtained for the Classical-Avoidance × Sex interaction in extinction ($F = 5.02$, $df = 1/48$, $P < .05$); and examination of the mean percentage CRs revealed that under the avoidance procedure males gave a higher percentage of CRs whereas, under the yoked-classical procedure the females were superior. In the remaining sources of variation those of statistical significance were without psychological import (i.e., interactions involving the recording system source of variation).

REFERENCES

MOORE, J. W., & GORMEZANO, I.   Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, 62, 552–559.

SNEDECOR, G. W.   *Statistical methods.*   (5th ed.)   Ames. Iowa: Iowa State Coll. Press, 1959.

## EDITORIAL

This issue of the *Journal of Experimental Psychology* marks the end of 144 issues and 12 years of my editorship. During these 12 years there have been marked changes in experimental psychology—chiefly in the directions of more mathematical formulation and testing of theory, more elaborate (and more adequate) experimental designs and statistical analyses, more emphasis on the human subject, and more emphasis on the "higher mental processes" of the human subject—whether in learning and problem solving or in "information processing" performance. The *Journal* has attempted to respond to and reflect these changes, even encourage them, at the same time that it remained a sympathetic medium for more traditional problems and methods and for the wide range of content properly described as the experimental analysis of the mental and behavioral processes of the individual organism—especially man.

Provided that the problem of an experimental study fit within the content boundaries prescribed for the *Journal*, the criterion for acceptance of an article has been at all times the question whether it warranted space in the ever-more-crowded archives of our science. This is, of course, a multidimensional criterion, some di-

mensions being quite objective and some being quite subjective. Objective, or at least rational, dimensions were matters pertaining to the adequacy of the experimental design for the collection of data on the problem as stated, the adequacy and appropriateness of the measures extracted from the data and the statistical tests employed, and the logical relationship between the data exhibited and the conclusions drawn. Criticism and rejection on the basis of these characteristics of the experiment may be considered as the application of "internal" criteria, since the emphasis is on internal consistency. Here the question was usually formulated as "Is this a valid experiment?" Many times the answer has been "No." Proper control groups were not tested; the design confounded variables that made the results and conclusions trivial, rather than important; the chosen method of summarizing the data led to conclusions that did not hold up if the data were summarized in another equally appropriate, or more appropriate, way; etc.

The next step in the assessment of an article involved a judgment with respect to the confidence to be placed in the findings—confidence that the results of the experiment would be repeatable under the conditions de-

scribed. In editing the *Journal* there has been a strong reluctance to accept and publish results related to the principal concern of the research when those results were significant at the .05 level, whether by one- or two-tailed test! This has not implied a slavish worship of the .01 level or any other level, as some critics may have implied. Rather, it reflects a belief that it is the responsibility of the investigator in a science to reveal his effect in such a way that no reasonable man would be in a position to discredit the results by saying that they were the product of the way the ball bounced. At least, it was believed that such findings do not deserve a place in an archival journal, even though they may be proper fare for symposia, scientific meetings, and dittoed handouts. The *P* level of a finding which was the major purpose of the investigation (anyone can find a significant practice effect) is only one element in the persuasion, others being the relation of necessity between the predicted relationship and other previously or concurrently demonstrated effects, and the consistency of the relationship across a sequence of experiments. But an isolated finding, especially when embodied in a 2 × 2 design, at the .05 level or even the .01 level was frequently judged not sufficiently impressive to warrant archival publication. The same philosophy applied when negative results were submitted for publication, but here rejection frequently followed the decision that the investigator had not given the data an opportunity to disprove the null hypothesis, i.e., the sensitivity of the experiment was substandard for the type of investigation in question and was therefore not sufficient to persuade an expert in the area that the variable in question did not have an effect as great as other variables of known significant effect.

Even if a proffered experimental study passed the hurdle of design adequacy and judged repeatability, it still needed to pass another hurdle, and an increasing number of articles failed this third hurdle as we moved into the late 1950s and early 1960s. Increasingly, we applied a criterion of *substantiality* to experimental studies. By this was meant that the investigation should not merely identify the effect of a variable, but should move beyond that simple demonstration to either the determination of a *function* relating levels of the variable to levels of the effect or the assembly of further information about the demonstrated effect—the composition of the variable, the range of tasks in which it had an effect, the variation of the effects as a function of a second or third variable, etc. We believed that the day of the archival report based on a simple experiment with an experimental and control group or with a 2 × 2 design was past in many mature areas of psychological research, and that each published report should make a more substantial contribution to the problem. In particular, it seemed desirable for experimental psychology to move toward the determination of quantitative functional relationships between independent and dependent variables, especially since so many of these quantitative relationships in behavior turn out to be nonmonotonic. Failure to make a serious effort to *understand* the variable, either through plotting the effects of several levels of it or through follow-up experiments with the intent of determining the generality or other contingencies of its effect, was considered sufficient reason to choose not to publish until such additional work had been done.

These, then, have been the guiding criteria for acceptance of articles for the *Journal*. When an article was rejected, an attempt was made to state the basis for rejection in terms of those criteria. But the same criteria were employed as the basis for required revisions prior to publication. Only about 10% of all articles received by the *Journal* were published without substantial revision or rejection, which is to say that four out of five of the published articles (which were, in turn, 50% of those received) have suffered substantial revision before publication. Very often the revision was required for the purpose of condensing the article, eliminating duplicated data in figures and tables, and in other ways decreasing the length of an article without reducing its essential content. But with a frequency that would surprise some readers, required revisions have consisted of adding detailed description of procedures, making explicit some design factor, adding data in tables or figures, and even urgings to the author to add words in order to make more explicit his analysis and interpretations. In short, the philosophy of acceptance has been that an article should not be rejected if the experiments were acceptable, even though major revision of the data analysis or the article as a whole was required. In fairness to contributors who submitted completely acceptable articles, those who were required to revise were given a limited period of time (usually 30 days) in which to resubmit the revision without loss of position in the publication order.

The intent of these criteria as applied either to acceptance or revision before publication has been to get more scientific mileage from the pages of the *Journal*. This was done by excluding questionable data, by encouraging the reporting of research in larger, more substantial chunks, and by reporting research as completely as necessary but also succinctly. All of this stems from the conception of the *Journal* as an archive of our science, not as a newssheet filled with a heavy load of transient, undigested, or fallible information. However, this screening of information for such archival records is not an infallible procedure. The criteria are not reducible to formula, and the final judgment is intrinsically subjective. Therefore, studies have been accepted and published, only later to be judged inadequate by the criteria; others have been rejected and published elsewhere, later to be widely acclaimed as containing important data. (I do not include in the latter class those experimental articles that deserved to be rejected because of grievous faults in the design, but which, when published elsewhere, stimulated research on a problem owing to the very inadequacies of the original experiment.)

It should be clear from what has been said about criteria that heavy demands were placed on the editorial staff for detailed information about what is already known and for methodological sophistication, and this demand applied to a myriad of special problem areas in the case of an omnibus journal such as the *Journal of Experimental Psychology*. The heart of the editorial system is now and has been the board of Consulting Editors of the *Journal*, some of whom served for the entire 12 years and all of whom have made multiple, essential contributions to the implementation of the criteria and standards that all of us considered to be in the best interest of the science. It is fitting, therefore, that all of the Consulting

Editors of the *Journal* be given this public repetition of my private expressions of appreciation and indebtedness for making my editorship of the *Journal* possible. The list follows:

Norman H. Anderson (1959–62)
E. James Archer (1957–62)
Fred Attneave, III (1959–62)
Judson S. Brown (1951–58)
Cletus J. Burke (1953–62)
James Deese (1960–62)
Paul M. Fitts (1957–62)
Frederick C. Frick (1957–59)
Robert M. Gagné (1953–56)
Wendell R. Garner (1954–56)
Frank A. Geldard (1951–62)
James J. Gibson (1951–62)
Clarence H. Graham (1951–61)
David A. Grant (1951–56)

Harold W. Hake (1957–62)
Lloyd G. Humphreys (1951–59)
Arthur L. Irion (1954–62)
Howard H. Kendler (1957–62)
Herschel W. Leibowitz (1962)
Donald B. Lindsley (1951–62)
Kenneth MacCorquodale (1957–62)
Quinn McNemar (1957–62)
Neal E. Miller (1951–57)
Edwin B. Newman (1951–54)
Leo Postman (1961–62)
L. Starling Reid (1958–62)
Kenneth W. Spence (1953–62)
Benton J. Underwood (1951–56)

Delos D. Wickens (1951–56, 1959–62)

---

In addition to those listed, many other psychologists have made important contributions to the *Journal* through their reviews of specific articles where their competence was deemed necessary for knowledgeable evaluations. Each of them, if he reads this, will, I hope, consider himself again privately thanked for his contribution.

This tribute to the Consulting Editors of the *Journal* must not be interpreted as shifting to their shoulders responsibility for the Type I and Type II errors we have made in accepting or rejecting articles. Their relationship to the final decision, which was always made by myself or by an Associate Editor, was understood at all times to be advisory. There were times—not very many— when Consulting Editors did not agree in their evaluation, and there were times—again, not very many— when we accepted even though the Consulting Editor recommended rejection, or rejected even though he recommended acceptance. In any event, he was informed of the action taken on his advice, since he received copies of the letters to authors which indicated acceptance, rejection, or requirements for revision. Perhaps we should take this opportunity to thank all those Consulting Editors who tolerated our bad judgment when we failed to follow their advice, and who did not resign forthwith—none did, to my knowledge.

Penultimately, I wish to express my deep appreciation to the three who served as Associate Editors of the *Journal*—David A. Grant (1957–62), Delos D. Wickens (1957–58), and William K. Estes (1959–62). Editing the *Journal* during these last 6 years would have been intolerable, if not impossible, without each of them assuming roughly one-third of the responsibility for deciding what should and should not be in the *Journal*. As many who have contributed to the *Journal* know, their roles were those of Co-Editors, with complete responsibility for conducting the relationship with authors up through the decision to accept or reject. The ability of the *Journal* to judge appropriately some of the technical and theoretical innovations of recent years is largely a consequence of their

participation in the editing of the *Journal*.

Finally, it would be thoughtless of me to bring this swan song to a close without expressing our appreciation of the tolerance (sometimes requiring incubation) that the vast majority of contributors to the *Journal* have shown toward the editorial mayhem performed, sometimes repeatedly, on the products of their thought and work. Experimental work in psychology is terribly laborious business, as every one who has done it knows. When, after all is done and a report of it is written, it is nothing short of mental cruelty to have an editor require that it be cut in half, and even worse to have him recommend that it serve as a lesson in how to do a better job next time. While I have no illusion that this editorial role has increased the quantity of warm sentiments that come my way, and I know that I have been hung in effigy in some laboratories and offices, it is still my hope that producing experimental psychologists who make our science grow apace, and who make the *Journal* possible, recognize the attempt to be fair and explicit, if not the wisdom, in the decisions they have suffered. Some authors have even been so kind as to say that this is so.

I feel no reluctance, only gratification and confidence, as I relinquish the editorship to my able friend and colleague, David A. Grant. These sentiments relate not only to his editorship, but also to the vigorous, sometimes combative, state of experimental psychology and experimental psychologists.

ARTHUR W. MELTON

# FREE RECALL LEARNING OF VISUAL FIGURES AS A FUNCTION OF FORM OF INTERNAL STRUCTURE

JAMES R. WHITMAN

*Veterans Administration Hospital, Perry Point, Maryland*

AND W. R. GARNER

*Johns Hopkins University*

The literature on factors affecting free recall learning is voluminous. The kinds of factors which have been investigated involve such things as the number of items to be learned, prior experience with the items, meaningfulness of the items, interitem similarity, etc. Most of the experiments reported have had as an assumption, either explicit or implicit, that characteristics of the individual items or stimuli are critical.

Garner (1962) has emphasized another aspect of free recall learning, namely, the internal structure inherent in groups of stimuli to be learned. A set of stimuli can be considered to be generated by a series of variables, and these variables can be interrelated in any subset of stimuli actually used in the learning experiment.[1] This relatedness constitutes the internal structure of the stimuli or the variables which make up the stimuli. The amount of internal structure is the same as the redundancy of the subset of stimuli and is determined by the number of stimuli used in an actual subset compared to the number which could have been generated with the same number (and levels) of the variables.

If the number of stimuli used in an experiment is the same as the total number which could be generated from the given variables, then all variables are orthogonal to each other and there is no internal structure to be learned. Knowing these variables, *S* can reproduce the set of stimuli without practice. But if the number of stimuli used is less then the number which could have been used, then internal structure exists; and both the amount and the form of this structure will affect ease of free recall learning.

The internal structure is not determined by the relations between elements of any particular stimulus. Rather, it is determined by the relations between the variables making up the stimuli across the subset of stimuli actually used. Thus the amount and form of internal structure cannot be specified without knowing the exact subset of stimuli used in the learning experiment.

These considerations led Garner (1962) to state two specific hypotheses, the tests of which are the purpose of this experiment. First, the ease of free recall learning is not a question of the characteristics of the individual stimuli which make up the subset but is a question of the characteristics of the entire subset of stimuli. Thus the same stimuli imbedded in two different subsets of stimuli will be learned according to the characteristics of the subset within which they are imbedded, and the nature of the unique stimuli is irrelevant.

[1] In this paper we shall use the term *set* to indicate the group of all possible stimuli generated by the specified variables and levels; we shall use the term *subset* for any group of stimuli which does not include all possible stimuli.

Second, the form of the internal structure is a critical factor in learning even with the same total amount of internal structure. Specifically, those forms of structure which involve direct contingencies between pairs of variables will produce easier learning than will forms of structure involving complex relations among three or more variables, i.e., interactions.

While these hypotheses are relevant to free recall learning of any stimuli, the present experiment tests them specifically with free recall learning of visual figures.

## METHOD

### The Stimuli

In carrying out an experiment to test the importance of the form of internal structure in free recall learning, it is important that the amount of internal structure be held constant. In more specific terms, it is important to specify not only the subsets of stimuli actually used but also the total set of stimuli which could have been used, since the ratio between these determines the amount of internal structure.

*Total set of potential stimuli.*—The potential stimuli, or the complete set of stimuli from which the subsets were selected, were formed by using three levels or values of each of four variables. If all possible combinations of these variables are generated, the total number of possible stimuli is 81. The four variables and their values are: (*a*) Shape, with squares, triangles, or circles constituting the levels; (*b*) Lines, with two, one, or zero lines bisecting the shape; (*c*) Spaces, with a space on the left, on the right, or no space; and (*d*) Dots, with a dot above the shape, below it, or none.

*Subsets of actual stimuli.*—Three different subsets of stimuli to be used in the experiment were chosen from this total set. Each subset contained 9 different stimuli from the 81 possible and differed only with regard to the form of the internal constraint. In selecting these three subsets, it is important that each subset demonstrate all four variables of the total set and, furthermore, that each level of each variable occur equally often. This precaution is necessary to ensure that the factor of total amount of internal structure not be confounded with the form of internal structure. All three subsets of actual stimuli



FIG. 1. The three subsets of actual stimuli used. (The number to the right of each figure provides the coded values for the four variables, in the order, shape, space, line, and dot. The underlined coded values for three figures each in Subsets B and C represent the three identical figures for these subsets.)

are shown in Fig. 1, along with coded values of the four variables.

Subset A: In the first subset of nine visual figures each of the four variables occurs three times at each level, but no two of the variables are directly correlated. Since there are four variables, there are six pairs of variables; and none of these pairs has a contingency greater than zero. This subset of stimuli is equivalent to a Graeco-Latin square, in which four variables are all orthogonal to each other.

Subset B: The second set of figures was selected so that one of the six pairs of variables was perfectly correlated and the other five were orthogonal. This subset provides a condition intermediate between subsets with minimum and maximum contingencies between variables.

Subset C: The last subset of figures was selected so that three of the six pairs of variables were perfectly correlated while the other three were uncorrelated. This subset provides the maximum contingency between pairs which can exist. Since nine different stimuli are required, at least one pair of variables must be orthogonal, a restriction which also means that no more than three of the pairs can be correlated.

TABLE 1

UNCERTAINTY ANALYSES OF THE FORM OF
INTERNAL STRUCTURE OF THE THREE
SUBSETS OF STIMULI, WITH SHAPE
(W), SPACE (X), LINE (Y), AND
DOT (Z) AS VARIABLES

| Contingencies | Subsets | | |
|---|---|---|---|
| | A | B | C |
| Simple | | | |
| W:X | 0 | 1.58 | 1.58 |
| W:Y | 0 | 0 | 1.58 |
| W:Z | 0 | 0 | 0 |
| X:Y | 0 | 0 | 1.58 |
| X:Z | 0 | 0 | 0 |
| Y:Z | 0 | 0 | 0 |
| Interaction | | | |
| WXY | 1.58 | 0 | −1.58 |
| WXZ | 1.58 | 0 | 0 |
| WYZ | 1.58 | 1.58 | 0 |
| XYZ | 1.58 | 1.58 | 0 |
| WXYZ | −3.16 | −1.58 | 0 |
| Total | | | |
| W:X:Y:Z | 3.16 | 3.16 | 3.16 |

Three of the stimuli in Subset C are identical to three in Subset B. These identical stimuli were included to allow comparison of learning rates for these particular stimuli, to determine the extent to which learning is affected by the particular stimuli rather than by the characteristics of the subset.

To summarize the characteristics of these three subsets of stimuli, each subset contains nine different stimuli; in addition, each subset contains exactly three occasions of each of the three levels of each of the four variables. Thus the number of specific elements which compose the different subsets is identical in all cases.

These subsets differ only in regard to the form of the internal structure, and an uncertainty analysis (in bits) of these three subsets of stimuli is shown in Table 1. The total amount of internal structure (W:X:Y:Z), shown at the bottom, is the same in all three cases, 3.16 bits. In Subset A, however, all of this structure is in the form of interactions, and none of the simple contingencies (between pairs of variables) is greater than zero. In Subset B, one simple contingency exists, and the rest of the structure is in the form of interactions. The pattern here is somewhat more complex. Again the interaction involving all four variables is negative and serves to correct the three-variable interactions. In Subset C, the

maximum amount of simple contingency exists since three of the pairs are perfectly correlated. Since this amount of structure is greater than the total structure, again the negative interaction term occurs to correct the total.

Our hypothesis concerning form of structure concerns the amount of structure which exists in the form of simple contingencies. In Subset A, none does; in Subset B, 1.58 bits does; and in Subset C, 4.74 bits does.

## Subjects

All $S$s were personnel associated with a large VA hospital and included 16 summer students and 9 staff members with professional degrees. They ranged in age from 15 to 58 yr. There was a total of 39 $S$s, and they were assigned randomly to each of three groups with the restriction that each group contain an equal number of professional staff and students, insofar as possible. (A median test showed no difference in performance between the different kinds of $S$s.) Each group of $S$s was required to learn just one of the three subsets of stimuli.

## Materials

The stimulus figures were drawn with black India ink on individual white pasteboard sheets, $8\frac{1}{2} \times 11$ in. The diameter of the circles and the sides of both triangles and squares were 6 in. All spaces were 2 in. in width and were centered. Dots were solid, $\frac{1}{4}$ in. in diameter, and centered $\frac{1}{2}$ in. below or above an edge. All lines were solid, about $\frac{1}{32}$ in. wide. Lines within the patterns were centered, and when two were present they were $\frac{1}{2}$ in. apart.

## Learning Trials

The $S$s were tested either individually or in small groups, depending on availability. The $E$ stood in front of $S$s and held each stimulus card from a subset for 5 sec., with one stimulus immediately following another. The order in which the stimuli were presented was predetermined so that no figure on any trial followed the same figure that it had on the preceding trial, and each figure was presented once as the first and once as the last in a series of nine trials.

The $S$s were told that they were participating in an experiment to see how fast they could learn to reproduce from memory nine different diagrams or figures which would be shown to them. The $E$ then described the four characteristics of the figures and the levels of each, giving illustrations by using

figures not later used in the experiment. The words "angles," "spaces," "lines," and "dots" were then written on a blackboard as a reminder of the four characteristics, and these remained in view throughout the experiment.

The Ss were then told that (a) the figures were going to be presented one at a time; (b) the order in which they were presented would vary; and (c) after they had seen all nine on a given trial they were to draw the figures from memory in any order. For this purpose they were given answer sheets containing nine blank spaces in three rows of three each. The Ss were instructed to draw nine different figures each time, guessing if necessary.

A trial consisted of the presentation of a complete subset of stimuli followed immediately by an answer period of 2 to 3 min. for Trial 1 and 1.5 to 2.5 min. for subsequent trials. At the end of 1.5 min. (2 min. for Trial 1) S was urged to complete nine different reproductions. At the end of the answer period, S covered his answers and was instructed not to look at them again. During Trials 1–5 E described each stimulus in terms of the four variables as it was presented. Practice continued for 20 trials or until S had correctly reproduced all nine figures on a single trial.

### Measures

The reproductions of S were scored in the following ways: (a) number of trials in order to reproduce correctly all nine figures; (b) number of correct responses on each trial; and (c) the amount of simple contingency between pairs of variables in the reproductions, without regard to correctness of response. This latter measure is simply a matter of determining the form of the internal structure in each set of nine reproductions in the same way that the stimuli themselves are described. In determining contingent uncertainties from the reproductions, an approximation procedure was used to facilitate computation. The number of pair coincidences was counted, and the total of these was translated into contingent uncertainty by a computed graphical function.

### RESULTS

*Form of structure.*—The main results pertain to the hypothesis concerning the effect of form of structure on free recall learning. Table 2 shows the number of trials required for a criterion of nine correct reproductions

for the three groups. Median trials were used, rather than means, because 6 of the 13 Ss learning Subset A had not learned the stimuli to the criterion within the 20 trials allowed. Subset A had stimuli with no pair contingencies, and it is clear that such a subset of stimuli is very difficult to learn. By contrast, Subset C, with maximum simple contingencies, was extremely easy to learn, with a median of just two trials. In fact, 5 of the 13 Ss learning Subset C correctly reproduced all nine stimuli on Trial 1.

Analysis of the data in terms of number of correct reproductions per trial shows equally clearly the great difference between these subsets of stimuli. This analysis, shown in Fig. 2, indicates how rapidly learning occurs with the high simple contingencies and how far from complete it is even after 20 trials with the zero contingencies (Subset A). The evidence in favor of the hypothesis could not be much clearer.

*Individual stimuli.*—Three stimuli in Subset B were identical to three stimuli in Subset C. If the characteristics of the individual stimuli are important in free recall learning, then these three stimuli should have been learned at the same rate regardless of the subset within which they were imbedded. Analysis of number of correct reproductions of just these three stimuli was carried out for each

TABLE 2

NUMBER OF LEARNING TRIALS TO CRITERION FOR THE THREE SUBSETS OF STIMULI

| Subsets | N | Range | Median | Mann-Whitney U Test |
|---------|---|-------|--------|---------------------|
| A | 13 | 9–20+ | 19 | A vs. B, 135.5* |
| B | 13 | 7–19 | 12 | B vs. C, 91.5** |
| C | 13 | 1–7 | 2 | |

\* P = .05.
\*\* P = .01.
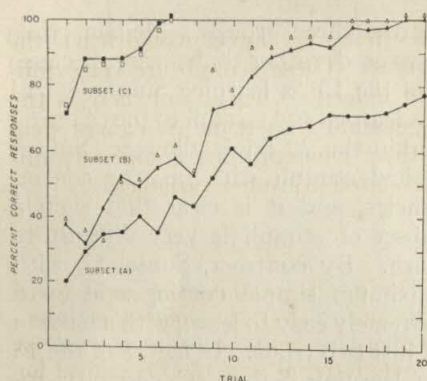
FIG. 2. Percentages of correct responses as a function of trial for the various subsets of figures. (The filled points are data for all nine figures of each subset. The open points are data from just those three figures in Subsets B and C which were identical. Each point is the mean for 13 Ss.)

subset separately, and the data obtained are plotted in Fig. 2 as the open squares and triangles. Since these data are plotted in terms of percentages of correct responses, direct comparisons of the learning curves are possible.

The curves for the three particular stimuli follow almost exactly the learning curves for the subsets within which they were contained and bear little relation to each other. The evidence could not be much clearer that the characteristics of the individual stimuli are of little relevance in free recall learning of subsets of stimuli but rather that the characteristics of the entire subset of stimuli are important.

In fact, when it is recalled that one-third of the stimuli in Subsets B and C were identical, the large difference in learning rates for these two subsets is even more impressive. Apparently what is learned is not the individual stimulus but a total set of relations between variables which make up the stimuli.

*Internal structure in reproductions.—* These data show that subsets of

stimuli in which there are high simple contingencies are easy to learn. In order to provide some additional understanding of the role of this factor in free recall learning, the amount of simple contingencies (in bits) was determined for Ss' reproductions on each successive trial; and these contingency results are shown in Fig. 3. There are several factors of interest in these curves.

First, in order to reproduce correctly all nine figures, the reproductions must contain the same pattern of contingencies as the stimuli themselves had. But this pattern is simply a prerequisite condition since it is possible to have the same total amount of simple contingencies but not to have the correct pairing of variables. Thus part of the learning process involves learning to reproduce the correct pattern of contingencies.

Second, analysis of the contingencies in the reproductions can give us some idea of what seems natural to Ss, and the data in Fig. 3 clearly show that Ss produce a very high level of
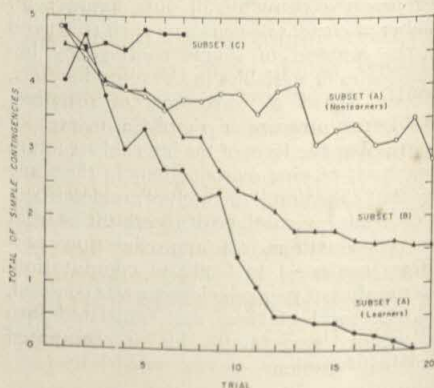


FIG. 3. Simple contingencies in reproduced subsets as a function of trial. (Each point is the average of the total amount of the simple contingencies in bits in the subsets as reproduced by S without regard to correctness of the reproductions. For Subsets B and C, each plotted point is the average of the 13 Ss' scores. The Ss for Subset A are divided into seven learners and six nonlearners.)

contingencies in their first few trials. This level, close to the maximum possible, is so high that there is actually very little learning for Subset C, in which the maximum possible contingencies are found. But for the other two subsets, this level is much too high, and in effect $Ss$ must learn to undo this apparently natural tendency to produce high contingencies before they can reproduce the stimuli correctly.

In order to obtain some idea of whether these high contingency values are simply the result of random pairings of variables, we produced 10 random sets of stimuli, in which it was only required that each level of each variable occur equally often and that all 9 patterns be different. The mean contingency value for these 10 randomly selected patterns was 3.23 bits, with a standard error of .18 bits. This value is so far below those for early trials that it is clear that $Ss$ do not just produce random amounts of simple contingencies but produce close to the maximum possible value. And these high values are produced even when the stimulus subsets themselves contained much lower values. Apparently these low values of simple contingency are contrary to $Ss'$ normal expectations.

The data for Subset A are plotted separately for $Ss$ who learned and those who did not learn the figures within the 20 trials. The nonlearners show very little evidence of learning at all, and there is the strong suggestion that some $Ss$ never would learn the figures. Actually, while the experiment was cut off at 20 trials, an attempt had been made to continue it for these nonlearners. Two of the six nonlearners were continued to 30 trials and had not yet learned. Two others refused to continue the experiment shortly after 20 trials, because they felt they never would learn the figures. These stimuli are very difficult indeed to learn, and there is the suggestion that some $Ss$ cannot deal with or conceptualize completely uncorrelated stimulus variables.

## DISCUSSION

The results of this experiment leave little doubt that the context of inter-relationships between variables within a subset (internal structure) is critical for free recall learning, and that the two hypotheses initially stated are valid: Free recall learning is a function of the structural characteristics of the entire subset of stimuli, not of the individual stimuli; and internal structure which exists in the form of simple contingencies between variables is better for free recall learning than are more complex forms of structure.

Each of these points deserves some comment, and we shall do so in reverse order. Miller (1958) showed that free recall learning is easier for what he called redundant strings of letters. He generated nonsense words by different statistical rules which affected the sequential dependencies between successive letters in the words and found that high sequential dependencies gave better learning. Since the lists which he compared, however, were of the same length, and since the number of different letters possible was the same for each list, it is clear that his experiment concerned not the amount of redundancy but rather its form. It is more difficult to state the amount of simple contingency in his lists since all words were not of the same length, but the nature of the differences was certainly similar to the differences used in the present experiment.

In this experiment, the *amount* of redundancy was the same in all three sets of stimuli, but the amount itself should be an important variable for free recall learning. Horowitz (1961) compared lists of letter trigrams differing in similarity by Underwood's (1954) definition of similarity as the extent to which words on the list share the same items. By this definition, low similarity is equivalent to

high redundancy or internal structure and vice versa since low similarity lists have many different levels or values per variable, a fact which means that the set of potential stimulus words is very high compared to the number of words actually used. He found better free recall learning in early trials for the high similarity (low redundancy) lists.

In generating his lists, Horowitz used, for the low redundancy lists, a form of redundancy in which pairs of variables (letter positions) were very nearly uncorrelated. As the present results show, such a form is poor for free recall learning. His high redundancy lists, on the other hand, had 12 different letters; and he used all of them in each of his three letter positions. Such a procedure means that each letter in each position is paired uniquely with a letter in each other position so that pair contingencies are necessarily high. With so many letters per position no other relation is possible. Yet the net effect is that Horowitz used a good form of structure with his high redundancy and a poor form for his low redundancy. It is almost certain that if Horowitz had used a good form of structure with his low redundancy lists, or even random pairings of letters, he would have obtained much larger differences between his low and his high redundancy lists.

There is another point that stems from Horowitz' experiment that needs emphasis here. He showed that the relations between similarity and learning depended on the kind of learning required. In the present context, it is almost certain that the results we have obtained are true for free recall learning, but they will probably not be true for kinds of learning which involve discrimination between items. Garner (1962) has discussed this problem in detail.

Many other experiments, summarized by Garner (1962), have shown that discrimination processes depend on the total set of stimuli rather than the individual stimuli. Klemmer and Loftus (1958), for example, showed that identification of numerals with brief visual exposures depends on the total set of forms within which the numerals are imbedded. Our experiment shows that similar considerations hold for learning processes.

It should be emphasized that the characteristics of a group of stimuli are not simply the sum of the characteristics of the individual stimuli but are characteristics which can exist and be specified only for the total subset. Thus what is learned is the entire subset. In actual fact, the problem must really be put in reverse: We cannot specify the characteristics of the individual stimulus until we know the characteristics of the entire subset since the nature of the required differentiations depends on the alternative stimuli within the subset.

## Summary

This experiment tested two hypotheses relating free recall learning to the form of the internal structure: (a) the ease of free recall learning depends not on the characteristics of the individual stimuli but on the characteristics of the entire subset to be learned; (b) when a subset of stimuli is characterized by simple contingencies between pairs of variables generating the set, free recall learning will be easier than when the subset is characterized by interactions involving three or more variables.

Three different forms of internal structure in subsets of visual figures were compared. The results showed clear differences in the predicted direction and both hypotheses were substantiated.

## References

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Horowitz, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, 62, 51–57.

Klemmer, E. T., & Loftus, J. P. Numerals, nonsense forms, and information. *USAF Cambridge Res. Cent. tech. Rep.*, 1958, No. 57-2. (ASTIA No. AD110063)

Miller, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, 56, 485–491.

Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, 61, 160–166.

# MUSCLE TENSION DURING MENTAL WORK UNDER SLEEP DEPRIVATION [1]

### ROBERT T. WILKINSON

### Applied Psychology Research Unit, Cambridge, England

It has been shown (Wilkinson, 1958) that there are some tasks which most people can perform as well as normally under 30 hr. sleep deprivation and also that there are some individuals whose performance seems quite unaffected by the stress whatever the task. Is lack of sleep completely without effect in these situations or is the effect appearing in some form which is not being measured? It has been suggested (Wilkinson, 1961) that motivational factors are important in deciding whether performance will be impaired; a man appears capable of performing normally in spite of loss of sleep if the rewards for doing so or the penalties for failing to do so are sufficiently great. The present hypothesis is that this will only be done at the expense of extra effort and that electromyographic (EMG) records of muscular responses may provide some indication of this. In this experiment, therefore, EMG has been measured concurrently with an assessment of the effect of loss of sleep on performance.

## METHOD

Procedure.—Twelve Ss, enlisted men between the ages of 18 and 30, carried out a 20-min. test of addition twice at an interval of 2–4 days, once with sleep and once without. The design was balanced for practice effects and for the possibility of the two test papers being of unequal difficulty. While doing the test, and for 2 min. before and after it, records of muscle tension (EMG) were taken.

The test.—Sitting alone in a cubicle Ss were given a sheet of 100 sums and required to complete as many as possible in 20 min. Each sum comprised five two-digit numbers to be added, the total to be written down and also spoken into a microphone. At the 15-min. point of the test E intervened, speaking to S through a loudspeaker in the cubicle. He said, "Now I want you to work faster and more accurately, and to help you I will tell you the time you take for each sum and whether you get it right or wrong." This knowledge of results (KR) was given throughout the last 5 min. of the test. On the previous day Ss were given a practice run, the procedure, including the recording of EMG being exactly the same as in the main tests except that the run lasted only 10 min. and E did not intervene with KR.

Sleep deprivation.—In their experimental test 6 Ss had been without sleep for some 56 hr. and the other 6 for about 32 hr. All were tested in the afternoon and this applied also to the control tests after normal sleep. The Ss carried out routine duties and some other tests while staying awake but were in no way overworked apart from the stress imposed by enforced wakefulness.

EMG recording.—EMG records were taken from a placement over the pronator teres muscle of the left (inactive) forearm, Ss being asked to allow the arm to hang loosely by their side as they sat at the table doing the sums or relaxing. A single-channel machine of private design was used having an input impedance of 250 KΩ; pulses reflecting the integrated output were recorded on one channel of a tape recorder while the other channel recorded by microphone the proceedings in the test cubicle. This record com-
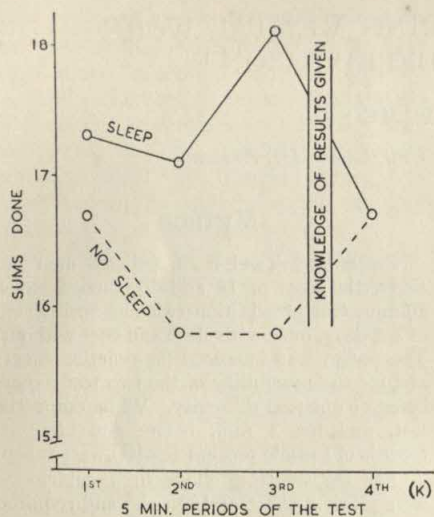
Fig. 1. Speed of adding with and without sleep.

prised mainly $S$'s answers to the sums and $E$'s encouragement and KR. Bipolar sponge electrodes were used and the skin was abraded to give reasonably equal and low resistance (between 10 KΩ and 3 KΩ) from each electrode to the reference electrode on the active forearm. As each $S$ was tested twice and comparisons made between the levels of EMG on each occasion it was essential that the placement and recording sensitivity should be approximately the same each time. To achieve this a patch of adhesive tape was placed over the proposed site of the electrodes. There were two holes in this patch, $1\frac{1}{2}$ in. apart and $\frac{3}{16}$ in. in diameter. In preparation for the first test the skin was abraded through these holes and the sponge electrodes placed immediately over them. The adhesive patch remained in place until the second test 2 or 4 days later when the electrodes were again placed over the holes and the skin abraded where necessary to achieve an electrode-to-electrode resistance stabilizing at approximately the same level as preceded the first test. Each test was preceded and followed by 2 min. relaxation when $S$ sat back in his chair and rested. EMG records were taken throughout and the score of "level of EMG" is the ratio of the average EMG during the test to the average during the preliminary period of relaxation. A further index of EMG is that of its variability in any given $S$ during a test. This score of EMG variability reflects the variance (calculated as the coefficient of variability) of the minute to

minute counts of EMG in each of the four 5-min. periods of the test.

*Statistical treatment.*—Significance of single means were tested by Wilcoxon's matched-pairs signed-ranks test, and differences between means by the Mann-Whitney $U$ test. Kendall's rank correlation coefficient ($\tau$) was used for all correlations. All these procedures are described by Siegel (1956). All significance levels refer to two-tail assessments except where otherwise stated.

## RESULTS

This section will give the results and their immediate implications; in the following section more general implications will be considered. The analysis that follows will be concerned mainly with the period of No KR (the first 15 min. of the test), but in addition we shall consider the changes that occurred when this feedback was added in the last 5 min. Finally, attention will be drawn to a possible predictor of the degree to which individual performance will be impaired under sleep deprivation.

*Period of No KR.*—During the first 15 min. of No KR sleep deprivation had no effect upon errors but it reduced the number of sums done (Fig. 1). This result was significant ($P < .01$) when the Practice × Order interactions were corrected for as follows: half the $S$s carried out their

TABLE 1

EMG LEVEL AND EMG VARIABILITY
WITH AND WITHOUT SLEEP

| 5-Min. Test Periods | EMG Level[a] | | EMG Variability[b] | |
|---|---|---|---|---|
| | Sleep | No Sleep | Sleep | No Sleep |
| 1 (No KR) | 1.88 | 1.95 | .119 | .227 |
| 2 (No KR) | 1.77 | 1.56 | .205 | .222 |
| 3 (No KR) | 1.63 | 1.82 | .137 | .273 |
| 4 (KR) | 2.88 | 2.01 | .244 | .317 |

[a] EMG level is the average EMG count during a given 5-min. test period divided by the average EMG during the preliminary 2-min. relaxation.
[b] EMG variability is the coefficient of variability ($V$) of the minute to minute counts of EMG.

first test under sleep deprivation and the second with normal sleep; for the other half the order of the conditions was reversed. All improved with practice from first to second test and the null hypothesis was that if sleep deprivation had no effect the practice effects of the two groups of Ss would not differ. Lack of sleep had little effect upon the level of EMG but it increased EMG variability ($P < .02$). These trends are shown in Table 1.

To examine concurrent trends of performance and EMG, Ss were ranked in order of impaired performance and of increased EMG due to loss of sleep, and the correlation of the two rankings was assessed. This operation had to be performed separately on each of the four teams of 3 Ss treated alike with respect to order and degree of sleep deprivation. The combined significance of the correlations was then assessed on a permutational basis to give an overall level of significance over all 12 Ss. To explain this further there are six possible combinations of the rankings of two sets of three scores. We have four teams in each of which any of these six combinations may occur. Over all four teams there are then $6^4 = 1296$ possible combinations of rankings. If we emerge with a combination of rankings whose correlations are predominantly negative, for example $-1.0$, $-1.0$, $-0.33$, and $-0.33$, in the four teams we can calculate the number of combinations out of the whole 1296 which are as negative as this or more negative. There are 41 in this case. The one-tailed probability of a negative combination as great or greater than this is then $41/1296$ or $.031$.

The negative correlations which emerged were almost all significant. Increased level of EMG due to loss of sleep correlated negatively with im-
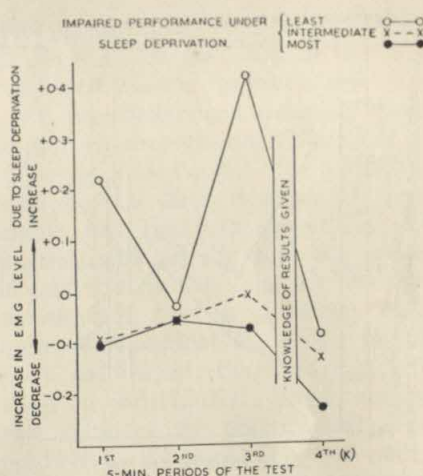


Fig. 2. Increase in EMG level due to sleep deprivation, i.e., $Log_{10}$ (No Sleep EMG level /Sleep EMG level) in three groups of Ss showing the least, the most, and an intermediate impairment of performance due to sleep deprivation.

paired performance in terms of both speed $(P = .031)$, and accuracy $(P = .094)$. Similarly increased variability of EMG under sleep deprivation correlated negatively with reduced speed $(P = .061)$ and reduced accuracy $(P = .007)$. Thus when sleep was lost those Ss whose performance was impaired least were the ones whose EMG was raised most and this holds good whether we correlate speed or accuracy of performance with either level or variability of EMG. To illustrate this (in terms of speed only) Ss have been divided into three groups containing the members of each team showing the least, the most, and an intermediate impairment of performance due to lack of sleep in the first 15 min. of the test. The tendency for these groups to show increased EMG as a result of losing sleep can be seen in terms of level of EMG in Fig. 2 and its variability in Fig. 3. Clearly there is an almost complete separation of the three per-
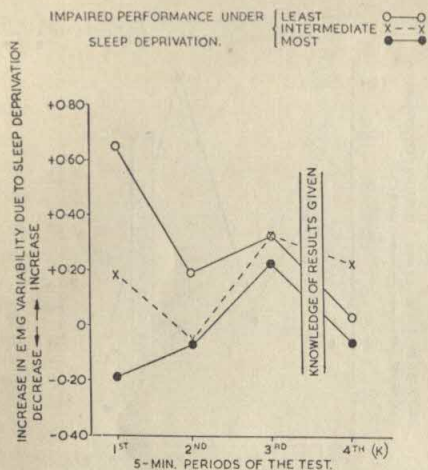
FIG. 3. Increase in EMG variability due to sleep deprivation, i.e., $Log_{10}$ (No Sleep EMG variability/Sleep EMG variability), in $Ss$ with the least, the most, and an intermediate impairment of performance due to sleep deprivation.

formance groups in the extent to which their muscle tension rose as a result of working without sleep.

These results seem very clear in this particular context, but they should be considered with due regard to the limitations of the experiment. They apply to only one form of activity. Only one physiological measure was taken, the EMG, and this was recorded from only one site. The conclusions which follow immediately and in later discussion should be regarded therefore as topics for confirmatory experiment rather than firm propositions. There are two immediate conclusions. The first is that although some men may be able to forego sleep and perform as well as normally on tasks of the present nature, this performance may be accompanied by abnormally high levels of muscle tension. Secondly, we may recall that Edwards (1941) concluded from incidental observation that work under sleep deprivation is accompanied by abnormally high expenditure of effort.

If we can assume that higher and more variable EMG is a sign of such effort the present result may provide more direct experimental evidence for this. A possible corollary is that sleep deprived men may be more uniformly inefficient than has been thought hitherto if we interpret efficiency in a mechanical sense as being the ratio of output to input. Previous implications have often been that where performance is maintained, so also is efficiency. This may only be true if effort remains the same also, and present results suggest that this is not always the case. Where output was maintained effort or input as judged by the EMG, was often higher. In such cases sleep deprivation may be reducing efficiency no less than when no extra effort is made and output falls.

*Period of KR.*—Errors showed no important changes as a result either of adding KR, or of sleep deprivation when this feedback was present. Performance is discussed therefore in terms of speed only.

When KR was given in the last 5 min. there ceased to be any difference between sleep deprived and normal performance (Fig. 1). Previous work (Wilkinson, 1961) has led us to expect this, but in the present experiment the result was brought about in an unusual way. When KR was added under sleep deprivation it raised EMG moderately and improved performance. When it was added after normal sleep, however, it raised EMG much more ($P < .01$) and this was accompanied by a *deterioration* in performance. These changes can be seen in Table 1 and Fig. 1. Now Stennett (1957) has shown that, beyond a certain point, increases in EMG may lower performance rather than improve it and it seems reasonable to account in this way for the

decline in performance among the sleepers when KR was added. They became overtense. In the circumstances it is not surprising to find that the negative correlation of the first 15 min. between impaired performance and increased EMG under sleep deprivation was reduced almost to zero when KR was added in the last 5 min. The lesson from this is that we should be careful not to generalize too far from results obtained in a relatively unstimulating situation (like the first 15 min. of the present test) to one in which incentives make S anxious to do well. In terms of efficiency as defined above the nonsleepers were no longer at a disadvantage when KR was added; indeed it could be argued that they were more efficient than the sleepers, for they performed as well and their EMG was lower. Clearly research of the present nature must be extended to more stimulating tasks.

*Prediction of individual impairment from EMG under normal conditions.*— Subjects may be ranked in terms of the ratio of their working level of EMG to that of their preliminary 2-min. period of relaxation, which, indeed, has been the index of level of EMG throughout. Three independent measures of this kind were obtained from each S, the first from the initial practice test and the second and third from the two main tests, one with and one without sleep. These three measures are in considerable agreement in their rankings of Ss, Kendall's coefficient of concordance being .67 and significant ($P < .03$). This suggests that the extent to which EMG rises in the transition from resting to working varies consistently from person to person. Table 2 summarizes the results of correlating these three assessments of this parameter with impairment of performance due to lack of sleep in the periods of No KR, of KR, and in the two combined, that is the whole test.

All the correlations involving speed of performance are negative, most of them significantly so. The main test after normal sleep is a reliable predictor, but the data concerning the practice test are the most interesting in that this measure was a truly predictive one, that is its results were quite independent of those to be predicted, namely the effect of lack of sleep on individuals. Unfortunately this value of the practice test as a predictive measure appeared only when all the results were analyzed. It was administered as no more than a practice run and with less care over EMG recording than was exercised in the main tests. But in spite of this it yields values of working-to-resting EMG ratio which predict in advance the impairment of speed of performance under sleep deprivation with fair accuracy and at nearly the .05 level of significance. The rankings also correlate ($\tau = .44$) ($P = .023$) with those of the highly predictive main test carried out under normal sleep. In short there seems good

TABLE 2

CORRELATIONS ($\tau$) OF THREE MEASURES OF WORKING-TO-RESTING RATIO OF EMG WITH THREE INDICES OF IMPAIRED PERFORMANCE (SPEED) UNDER SLEEP DEPRIVATION

| Working-to-Resting Ratio of EMG | Impaired Performance | | | | | |
|---|---|---|---|---|---|---|
| | No KR | | KR | | Whole Test No KR + KR | |
| | $\tau$ | $P$ | $\tau$ | $P$ | $\tau$ | $P$ |
| Main test (with sleep) | .55 | .006 | .53 | .008 | .67 | .001 |
| Main test (no sleep) | .39 | .04 | .35 | .06 | .39 | .04 |
| Practice | .24 | .13 | .33 | .06 | .36 | .051 |

Note.—All $\tau$ coefficients are negative.

reason to believe that if a careful preliminary assessment of the ratio of working-to-resting EMG is made this index should predict the degree to which performance will be impaired by lack of sleep in any subsequent performance of the task, the higher the ratio the less the impairment.

## DISCUSSION

Muscle tension (EMG) is one of a number of physiological measures which are sometimes (Malmo, 1959) thought to reflect the level of arousal of the body as defined by Duffy (1957), Lindsley (1951), and Hebb (1955). Other possible measures include pulse, respiration and metabolic rates, skin conductance, urinary excretion of catechol amines, and alpha depression in the EEG. When these are recorded under sleep deprivation their levels are sometimes higher than normal (Freeman, 1932; Hasselman, Schaff, & Metz, 1960; Laird & Wheeler, 1926; Malmo & Surwillo, 1960; Tyler, Goodman, & Rothman, 1947) and sometimes lower (Armington & Mitnick, 1959; Ax & Luby, 1961; Bjerner, 1949). Similar contrasts occurred with EMG in the present experiment. The fact that increased EMG under the stress correlated positively with maintained performance suggests that this, and perhaps other physiological indices may rise under sleep deprivation as the experimental situation is stimulating and provokes effort. If we examine the conditions under which physiological measures were taken in previous experiments the impression is reinforced; where levels increased the Ss were usually engaged in relatively stimulating tasks; where they fell the tasks appear less stimulating or else the Ss were merely sitting passively.

If these physiological indices reflect the level of arousal we must conclude with Malmo and Surwillo (1960) that sleep deprivation can either raise or lower arousal according to the situation in which the S is placed during recording. But do they? With No KR in the present experiment performance was impaired under sleep deprivation but the level of EMG was unchanged (Fig. 1 and Table 1); if this implies unchanged arousal the relationship between arousal and performance is broken. Similarly with KR different levels of EMG accompanied the same level of performance with and without sleep. Perhaps if we wish to retain the inverted U relationship between arousal and performance (Hebb, 1955) we must sacrifice the notion that EMG level always reflects the level of arousal. In particular when sleep has been lost it seems likely that higher levels of EMG are required for given levels of arousal. This suggests an explanation of the abnormally high levels of the so-called arousal measures which occurred under certain circumstances in the present and other experiments: they may reflect, not raised arousal, but the effort associated with maintaining normal arousal and customary standards of performance in face of the influence of sleep deprivation per se which may be always towards lowered arousal.

## SUMMARY

Twelve Ss performed a 20-min. test of addition, once after normal sleep and once under 32–56 hr. sleep deprivation. Records of muscle tension (EMG) were taken from the inactive arm. The Ss who maintained performance best under the stress showed the greatest rise in EMG over normal levels. Knowledge of results disturbed this relationship. An independent measure of EMG taken under normal conditions predicted those Ss whose performance was impaired. Sleep deprivation may cause inefficiency even in Ss who maintain performance if their raised EMG reflects greater effort or energy expenditure; this may be the cost of maintaining normal levels of arousal and performance in face of the depressing influence of sleep deprivation per se.

## REFERENCES

ARMINGTON, J. C., & MITNICK, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.*, 1959, **14**, 247–250.

Ax, A., & LUBY, E. D. Autonomic responses to sleep deprivation. *Arch. gen. Psychiat.*, 1961, **4**, 55–59.

BJERNER, B. Alpha depression and lowered pulse rate during delayed actions in a serial

reaction test. *Acta physiol. Scand., Stockholm*, 1949, **19**(Suppl. No. 65).

DUFFY, E. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, **64**, 265–275.

EDWARDS, A. S. Effects of the loss of 100 hours of sleep. *Amer. J. Psychol.*, 1941, **54**, 80–91.

FREEMAN, G. L. Compensatory reinforcements of muscular tension subsequent to sleep loss. *J. exp. Psychol.*, 1932, **15**, 267–283.

HASSELMAN, M., SCHAFF, G., & METZ, B. Respective influences of work, ambient temperature and sleep deprivation on the urinary excretion of catechol amines in the normal man. *CR Soc. Biol., Paris*, 1960, **154**, 197–201.

HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

LAIRD, D. A., & WHEELER, W. What it costs to lose sleep. *Industr. Psychol.*, 1926, **1**, 694–696.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

MALMO, R. B., & SURWILLO, W. W. Sleep deprivation: Changes in performance and physiological indicants of activation. *Psychol. Monogr.*, 1960, **74**(15, Whole No. 502).

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

STENNETT, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

TYLER, D. B., GOODMAN, J., & ROTHMAN, T. The effect of experimental insomnia on the rate of potential changes in the brain. *Amer. J. Physiol.*, 1947, **149**, 185–193.

WILKINSON, R. T. The effects of sleep loss on performance. *Med. Res. Council, Appl. Psychol. Res. Unit Rep., Great Britain*, 1958, No. 323.

WILKINSON, R. T. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *J. exp. Psychol.*, 1961, **62**, 263–271.

# THE PARASITIC REINFORCEMENT OF VERBAL ASSOCIATIVE RESPONSES [1]

W. D. KINCAID, JR.,[2] W. A. BOUSFIELD, AND G. A. WHITMARSH [3]

*University of Connecticut*

In recent papers Bousfield, Cohen, and Whitmarsh (1958b), and Bousfield, Whitmarsh, and Danick (1958) have attempted to account for the phenomenon of verbal stimulus generalization on the basis of the overlap of verbal associative responses elicited by the given words. Their basic approach rests upon the assumption that the presentation of a meaningful stimulus word leads to the elicitation of a composite of implicit verbal associative responses. They reason that the verbal conditioning of a response to a stimulus word involves not only the conditioning to the stimulus word, but also a simultaneous conditioning to the composite of verbal associative responses to that word. Thus, during conditioning trials, members of the associative response composite are involved in the learning process through higher-order conditioning. The term *parasitic reinforcement* may be used to describe this concurrent conditioning of the members of the composite of verbal associative responses. This term was introduced by Morgan and Underwood (1950) to explain the following phenomenon. After the learning of a given verbal response, B, to a stimulus word, A, the synonyms of B will have a greater than chance probability of subsequently being elicited by A. Bousfield, Whitmarsh, and Danick (1958) extended the concept of parasitic reinforcement to include the aggregate of verbal associative responses to a given stimulus word. Support for the assumption underlying the concept of parasitic reinforcement was found in the fact that the degree to which an observable response which had been conditioned to one word was elicited by the presentation of a second word was a function of the verbal associative responses common to the two stimulus words. Studies by Cohen (1958) and by Whitmarsh and Bousfield (1961) have replicated these findings, and have shown them to be independent of a specific technique used to measure generalization. While the theoretical rationale introduced to account for the contribution of associative responses to generalization assumed the conditioning of the implicit verbal associates of the first stimulus word to the observable response, these studies provide only indirect support for this assumption.

The present study was undertaken to test the deduction that after paired-associate learning the associates of the learned response word may also be elicited by the stimulus item of the learned pair. Specifically we wished to test the following hypothesis: the paired-associate learning of a meaningful response word to a nonsense-syllable stimulus has the consequence of establishing connections between the nonsense syllable and the members of a group of verbal associative re-

[2] Now at Androscoggin Mental Health Clinic, Lewiston, Maine.

[3] Now at the Springfield State Hospital, Sykesville, Maryland.

## TABLE 1
MEANINGFUL RESPONSE WORDS USED IN TRAINING
AND THEIR TESTED ASSOCIATES: EXP. I

| Words Used as Responses in Training | Associates of the Learned Responses and their Cultural Frequencies of Occurrence | | | | | |
|---|---|---|---|---|---|---|
| | High | | Medium | | Low | |
| ANIMAL | Dog | 81 | Cat | 17 | Man | 5 |
| ICE | Cold | 93 | Water | 13 | Cream | 12 |
| LETTUCE | Tomato | 38 | Green | 18 | Leaf | 11 |
| MOSQUITO | Bite | 67 | Bug | 26 | Insect | 13 |
| PETAL | Flower | 39 | Rose | 16 | Leaf | 11 |
| RAYON | Silk | 36 | Nylon | 25 | Material | 7 |
| TABLE | Chair | 45 | Write | 9 | Office | 5 |
| TIN | Can | 72 | Metal | 29 | Roof | 9 |
| TYPHOID | Fever | 74 | Disease | 22 | Sickness | 7 |
| WAGON | Wheels | 76 | Train | 12 | Red | 4 |
| Mean | | 62.1 | | 18.7 | | 8.4 |

Note.—See text for definition of high, medium, and low grouping.

sponses to the learned response word. For example, if the word RAYON were learned as a response to the nonsense syllable GOX, we should expect to find evidence of acquired connections between GOX and the associative responses to RAYON, as for example, Silk, Nylon, Material, and Soft. The relative strengths of the members of the composite of verbal associative responses to a given word may be measured from their cultural frequencies of occurrence as responses to that word in free associational norms of the Minnesota type (Russell & Jenkins, 1954). Our second experimental hypothesis concerns the relation between cultural habit strengths of the associates of a given word and their susceptibility to conditioning: the strength of the connections established between the nonsense syllable and the associates of the learned response word is an increasing function of the cultural habit strengths of the associates as responses to the learned response word.

## EXPERIMENT I

### Method

The materials for the initial learning were 10 pairs of nonsense syllables and meaningful words. The following 10 nonsense syllables were selected from the Glaze (1928) list on the basis of their having association values ranging from 0 to 47%: GOX, HAJ, MUP, NID, QOL, RUC, SIW, VEK, YEF, and ZAB. The 10 meaningful words, which are listed in Table 1, were selected from a list of 150 words for which free-associational norms had been compiled from a population of 150 Ss. Three different randomized pairings of these items were then prepared so that no nonsense syllable was paired with the same word more than once. The items for the testing phase of the experiment were free associational responses to the 10 response members of the initial learning pairs selected on the following basis. The 150 associational responses to each of the 10 learned words were divided into tertiles on the basis of their cultural frequencies of occurrence in the normative data. The associates in each of these three groups then comprised the pools of low-, medium-, and high-frequency associates. In choosing associates for the testing phase of the experiment the restriction was imposed that a chosen associate to a given word should not appear in the gradient of associational responses to any of the other 9 learned words. Within this restriction, the associate having the highest frequency in each frequency group was chosen as one of the three test words for a given learned word. The three associates thus chosen for each of the 10 learned words are listed in Table 1 along with their corresponding cultural frequencies of occurrence as associates to the learned word. It may be noted that Leaf appears as a low-frequency associate to both LETTUCE and

PETAL. This violation of the restriction imposed in the selection of associates was necessitated by the relative lack of degrees of freedom imposed by the limited pool of 150 stimulus words.

The Ss were 140 undergraduate students who were trained and tested in three groups comprising 51, 44, and 45 Ss, respectively. Each group received a different randomization of the paired-associate lists for learning. The items were presented one at a time to Ss by means of a Selectroslide projector set for exposures of 2.5 sec. The instructions and procedure employed for the paired-associate learning were the same as those devised by Cohen (1958) for group-method experiments. Eight learning trials were administered to all Ss. On alternate trials Ss were asked to anticipate and write in a booklet provided for this purpose the response member of the pair when shown the nonsense syllable. They were then verbally presented with the correct word. A total of 27 Ss failed to reach the criterion of all correct anticipations on the last trial and were therefore dropped. This resulted in Ns of 43, 33, and 37 Ss, respectively, for the three experimental groups.

After the initial learning, E proceeded immediately to the testing phase. A pilot study had demonstrated that the procedure of simply presenting the nonsense syllable with instructions for free association was effective in eliciting associates to the learned words in only 40% of the cases. This consideration led to the development of an alternative procedure. The S was given 10 data sheets, in booklet form, one sheet for each of the nonsense syllables of the training phase of the experiment. The nonsense syllable was followed by five words, one of which was one of the three chosen associates of the learned word. For example, for Ss who learned the pair GOX-RAYON, one of the five words was Material, a low-frequency associate of RAYON. The remaining four control words were selected at random from a dictionary with the restriction that they should not appear as an associate of any one of the 10 words used in the learning. The S was instructed to check the one word which he felt to be "most related to the nonsense syllable." [4] The order of the syllables in the booklet was randomized between Ss as was the position of the critical

associate among the four control words. Each group of Ss received booklets containing high-, medium-, and low-frequency associates distributed among the 10 nonsense syllables and among the three groups of Ss in a counterbalanced design. This design required the use of 120 control words.

With the forced-choice test instructions it might be supposed that the probability of selecting any one of the five choices would be .2. Such an assumption, however, appeared unsafe in view of the possibility that an alternative might be selected on the basis of extraneous factors such as phonetographic similarity to the nonsense syllable. It appeared advisable, therefore, to obtain what may be called base frequency data from a group of control Ss. For this purpose three control groups of undergraduates comprising 51, 33, and 34 Ss, respectively, were presented the test booklets and the same instructions as were given to the three groups of experimental Ss. Thus, each of the 30 forced-choice association tests, i.e., 10 for the low-, 10 for the medium-, and 10 for the high-frequency responses, was taken by 51, 33, or 34 Ss.

## Results

The first step in the treatment of the data was that of tabulating the total frequency with which each of the 30 associates used in the testing was selected as most related to its associated syllable. For example, the pair MUP-MOSQUITO appeared in the initial learning. In the testing situation the group of 43 Ss who had received this pair for learning was presented with MUP and asked to select the word most related to MUP from five alternatives comprising Insect, the low frequency associate of MOSQUITO, and the control words Knife, Field, Crazy, and Word. In view of the predicted facilitation of the associative responses to MOSQUITO, the checking of Insect as the preferred alternative was for convenience labeled "correct." Control word choices were designated as "incorrect." In these terms all 43 of the experimental Ss who were presented with this set of choices gave "correct" responses, whereas 17 of the 51 control Ss who received the same test gave the

---

[4] In a subsequent study these instructions were rephrased so that Ss were asked to check the one word which the stimulus item "most makes you think of." There was no evidence to indicate the choices of Ss were altered by this change in the instructions.

TABLE 2

PROPORTION OF EXPERIMENTAL (E) AND CONTROL (C) $Ss$ SELECTING EACH ASSOCIATE AND THE DIFFERENCES (E − C) BETWEEN THESE PROPORTIONS: EXP. I

| Word | High | | | Medium | | | Low | | |
|------|------|------|-------|--------|------|-------|------|------|-------|
|      | E    | C    | E − C | E      | C    | E − C | E    | C    | E − C |
| ANIMAL   | .953  | .196 | .757 | .973  | .238 | .735 | .939  | .303 | .636 |
| ICE      | 1.000 | .157 | .843 | .784  | .206 | .578 | .758  | .152 | .606 |
| LETTUCE  | .891  | .238 | .653 | .970  | .273 | .697 | .907  | .118 | .789 |
| MOSQUITO | .865  | .206 | .659 | .939  | .121 | .818 | 1.000 | .333 | .667 |
| PETAL    | .953  | .294 | .659 | .838  | .176 | .662 | .818  | .121 | .697 |
| RAYON    | .865  | .118 | .747 | .970  | .121 | .849 | .977  | .235 | .742 |
| TABLE    | .838  | .176 | .662 | .424  | .121 | .303 | .814  | .160 | .654 |
| TIN      | .730  | .059 | .671 | .953  | .260 | .693 | .909  | .182 | .727 |
| TYPHOID  | .909  | .212 | .697 | 1.000 | .314 | .686 | .973  | .294 | .679 |
| WAGON    | .939  | .303 | .636 | .953  | .140 | .813 | .838  | .235 | .603 |
| Mean     |       |      | .698 |       |      | .683 |       |      | .680 |

Note.—All E − C values are significant in the predicted direction at less than the .01 level.

so-called "correct" responses and 34 gave responses labeled "incorrect." A chi square analysis of these data with a two-tailed test indicates that the difference between experimental and control group responses is significant beyond the .01 level in the direction predicted by the experimental hypothesis. A similar treatment of the experimental and control group data for the remaining 29 associates indicated that all differences were significant beyond the .01 level in the predicted direction.

The next step taken in the analysis of the data was that of determining the nature of the relationship between the cultural frequencies of the associates as represented in the three groups of high, medium, and low on the one hand, and the extent to which these associates were facilitated in the testing phase of the experiment. The mean cultural frequencies of these associates, based on the normative population of 150 $Ss$, were, respectively, 62.1, 18.7, and 8.4. The prediction was that the number of responses labeled correct by the experimental $Ss$ should be greatest for the high-

frequency associates and least for those of low frequencies. The following steps were taken in this analysis. First, the proportion of $Ss$ who gave the so-called correct responses was determined for each of the 30 associates listed earlier in Table 1. These proportions appear in Table 2, and are listed in Column E for the experimental $Ss$ and in Column C for the control $Ss$ who supplied the base frequency data. Thus, the high-, medium-, and low-frequency associates of ANIMAL were, respectively, Dog, Cat, and Man. Table 2, Column E, shows that the proportion of experimental $Ss$ who checked Dog as related to the nonsense syllable which had been paired previously with ANIMAL was .953. The proportion of control $Ss$, Column C, who selected Dog was .196. The difference between these proportions, .757, is listed in Column E − C. This difference may be said to represent the effect of learning. As indicated earlier, this difference is significant. The means of these adjusted proportions for the high, medium, and low associates are, respectively, .698, .683, and .680. Three $CR$

tests of the differences between the proportions were performed for these three adjusted means. The differences between the means of the adjusted proportions for the high vs. medium, medium vs. low, and high vs. low groups are .015, .003, and .018, respectively. These mean differences do not differ significantly. Thus, while the findings of Exp. I support the first hypothesis, the variation in strength of the associates of the response words used in the initial training did not prove to be a significant parameter as predicted in the second hypothesis.

## EXPERIMENT II

In light of the unexpectedly strong effects of the so-called low-frequency associates in Exp. I, Exp. II was undertaken to extend the range of cultural frequencies tested to associates occurring only once in the normative population of 150 Ss.

The same nonsense-syllable meaningful-word pairs used in Exp. I were learned by the

### TABLE 3
MEANINGFUL RESPONSE WORDS USED IN TRAINING AND THEIR TESTED ASSOCIATES: EXP. II

| Words Used as Responses in Training | Associates of the Learned Responses and Their Cultural Frequencies of Occurrence | | |
|---|---|---|---|
| | Low | Low-Low A | Low-Low B |
| ANIMAL | Bear 3 | Ugly[a] | Human |
| ICE | Berg 3 | Hard | Winter |
| LETTUCE | Money 4 | Potato | Chow |
| MOSQUITO | Gnat 3 | Pest | Nasty |
| PETAL | Push 4 | Fall | Brake[a] |
| RAYON | Soft 3 | Skirt | Yarn |
| TABLE | Paper 4 | Book | Brown |
| TIN | Copper 4 | Rubber[a] | Pail |
| TYPHOID | Illness 3 | Gear[a] | Neck |
| WAGON | Red 4 | Children | Drunk[a] |
| Mean | 3.5 | 1 | 1 |

[a] Indicates the five associates which experimental Ss did not select with frequencies significantly different from the choices indicated in the base frequency data.

136 undergraduate Ss participating in Exp. II. The Ss were trained and tested in three groups of 43, 49, and 44 Ss, respectively. The critical associates tested in Exp. II were divided into three groups: one of 10 associates with cultural frequencies of either 3 or 4 which is designated the Low group, and two groups of 10 associates, each associate having a cultural frequency of 1. These groups were designated, respectively, Low-Low A and Low-Low B. The associates having the frequency of 1 were selected at random from the normative data with the restriction that they did not appear among the free associates to the other 19 words in the two Low-Low lists. The same forced-choice test procedure, counterbalanced experimental design, and instructions used in Exp. I were employed. The items used in Exp. II are presented in Table 3. A total of 27 Ss served as controls and provided the base frequency normative data used in this experiment. An analysis of the normative data collected in Exp. I indicated that adequately stable data could be obtained with an N of this size.

## Results

In Table 4, Column E shows the proportion of experimental Ss who gave the so-called correct responses for each of the 30 associates used in this experiment. Column C lists the proportions of "correct" responses for the control Ss, and Column E−C shows the differences between the experimental and control proportions.

Individual chi square tests were performed on the differences between the "correct" choices of the experimental Ss and the control Ss for each of the 30 associates. This analysis indicated that with one exception all associates in the low-frequency group were chosen by Ss at significance values beyond the .01 level in the direction predicted by the experimental hypothesis. The associate of PETAL, namely, Push, was significant at the .05 level. Thus, the results strongly confirm the first experimental hypothesis even when the cultural frequencies of the associates tested are further reduced in magnitude as com-

TABLE 4

PROPORTION OF EXPERIMENTAL (E) AND CONTROL (C) Ss SELECTING EACH ASSOCIATE AND THE DIFFERENCES (E − C) BETWEEN THESE PROPORTIONS: EXP. II

| Word | Low | | | Low-Low A | | | Low-Low B | | |
|---|---|---|---|---|---|---|---|---|---|
| | E | C | E − C | E | C | E − C | E | C | E − C |
| ANIMAL | .909 | .074 | .835** | .302 | .593 | −.291 | .796 | .074 | .722** |
| ICE | .864 | .296 | .568** | .605 | .333 | .272** | .898 | .074 | .824** |
| LETTUCE | .721 | .111 | .610** | .653 | .111 | .542** | .932 | .222 | .710** |
| MOSQUITO | .861 | .074 | .787** | .939 | .370 | .569** | .568 | .148 | .420** |
| PETAL | .341 | .185 | .156* | .395 | .185 | .210** | .408 | .370 | −.038 |
| RAYON | .861 | .259 | .602** | .939 | .074 | .865** | .705 | .222 | .483** |
| TABLE | .302 | .111 | .191** | .653 | .111 | .542** | .932 | .037 | .895** |
| TIN | 1.000 | .370 | .630** | .159 | .111 | .048 | .930 | .222 | .708** |
| TYPHOID | .837 | .296 | .541** | .000 | .185 | −.185 | .186 | .037 | .149** |
| WAGON | .673 | .185 | .488** | .636 | .037 | .599** | .093 | .296 | −.203 |
| Mean | | .541 | | | .317a | | | .467a | |

a Mean E − C of Low-Low A and Low-Low B combined = .392.
* .01 < P < .05.
** P < .01.

pared to the low-frequency associates of Exp. I. Similarly, chi square tests were made on the 20 Low-Low associates having cultural frequencies of occurrence of 1. This analysis indicated that 15 of these associates were selected by the experimental Ss at or beyond the .05 level of significance when compared with the base frequency data by means of two-tailed chi square tests. The five associates which did not attain significance are indicated in Table 3. Thus, connections were established between the nonsense syllables and 75% of the Low-Low associates of the learned response words even when the cultural frequencies of these associates as responses to the learned words were so low as to occur only once in a normative group of 150 Ss.

Several comparisons were made between the data provided by the two experiments. The means of the E − C proportions of the two Low-Low groups used in Exp. II were combined after a CR test indicated that the difference between these two groups was not significant. Two-tailed CR tests for uncorrelated proportions were made on the differences between all frequency groups in Exp. I and those in Exp. II and between the low-frequency group and the combined Low-Low groups of Exp. II. No significant differences between any of these adjusted mean proportions were obtained. Although there is a trend of decreasing mean proportions of correct responses for the frequency groups used in both experiments, the statistical analyses indicated that none of the means involved in this trend showed significant differences between each other. Even the difference between the data for the High group of Exp. I and the combined Low-Low groups of Exp. II was not significant. The second hypothesis was not supported.

DISCUSSION

The findings support the assumption that the learning of a meaningful verbal response to a nonsense syllable stimulus results in the establishment of measurable associative relationships between the stimulus and the verbal associative

responses to the meaningful word. Two alternative explanations of this phenomenon may be considered. The effect may be a consequence of mediation in the testing phase of the experiment provided by recall of the learned response to the nonsense syllable. The $S$ may recall that he has learned, for example, RAYON as the response to GOX. The associates of RAYON, namely, Soft, Material, etc., are then mediated by the recall of RAYON, and $S$ proceeds to check the associate Soft as the response most related to GOX. On the other hand the theoretical approach of Bousfield and his associates suggests that the phenomenon is attributable to the higher-order conditioning of the implicit verbal associates of the learned response word during the training phase of the experiment. A test of the assumption of the training phase locus of the effect would require the demonstration of parasitic reinforcement of the associative responses when the learned response had been forgotten and was no longer available to $S$. Failure to demonstrate the phenomenon under this condition, however, would not necessarily indicate that the locus was in the testing phase since it may very well be that the time interval necessary for the forgetting of the originally learned response word is also sufficient for the forgetting of the associational responses. While the locus of the effect has not been tested directly, some support for a training phase locus may be found in a study by Yavuz and Bousfield (1959) who showed that the connotative meaning of a foreign word could be recalled after the supposed English translation of the word had been forgotten. They suggested that the conditioning of the associational responses of the English word in the training phase mediated the meaning judgments of $S$s.

It would seem that the failure to find a differential effect as a function of the habit strengths of the associative responses may be attributed to either of two factors or to a combination of these factors. In the first place, it is evident that the findings here reported derive in part from the use of a particular method

for appraising the presence of the associative connections assumed to have been established in the initial learning. In each of a series of tests, the $S$s were given one of the nonsense syllables encountered in the initial learning which was followed by five different words. The $S$s were told to choose the one word of these five which they judged to be most related to the given nonsense syllable. In each case one of the five alternatives was an associate of the word learned as a response to the nonsense syllable. According to the theory outlined by Bousfield et al. (1958), this associate should have been elicited implicitly in contiguity with the presentation of the nonsense syllable during learning. It may therefore be said that the testing method actually employs the method of recognition. This method typically yields relatively high scores in tests of retention as long as the learned items are embedded in dissimilar new items as was the case in the present study (Luh, 1922). The sensitivity of the method of recognition is most likely due to the opportunity it provides $S$ for making use of relatively weak associations.

An alternative explanation of the failure of the findings to discriminate between the strengths of the associative habits is possible. It is conceivable that associative response strengths represented by a cultural frequency of 1 in a population of 150 are of sufficient potency in certain situations to become as effective as the associative responses whose strengths are reflected in higher frequencies of occurrence. If this is so, it would suggest that more attention needs to be paid to the so-called weak associative habits in the study of verbal behavior. Perhaps these habits are not as weak in effect as might be supposed from their cultural frequencies of occurrence. A similar phenomenon has been found in several studies employing Thorndike-Lorge frequency of usage values in which differences in performance as a function of high- or low-frequency values, while significant, are small in absolute differences (Bousfield, Cohen, & Whitmarsh, 1958a; Hall, 1954). In discussing this,

Underwood and Schulz (1960) suggest that "even words with the lowest frequencies may in fact have been experienced many times by a subject who serves in learning experiments" (p. 59). It may be that in studies of this type a distinction must be made between relative and absolute differences.

## SUMMARY

Two experiments were designed to test the following hypothesis: (a) The paired-associate learning of a meaningful response word to a nonsense syllable stimulus has the consequence of establishing connections between the nonsense syllable and the members of a group of verbal associative responses to the learned response word. (b) The strength of the connections established between the nonsense syllable and the associates of the learned response word is an increasing function of the cultural habit strengths of the associates as responses to the learned response word.

Ten nonsense-syllable meaningful-word pairs were presented for paired-associate learning to 113 Ss in Exp. I and to 136 Ss in Exp. II. The response words were selected from a set of stimulus words for which free associational norms had been previously obtained from a population of 150 Ss. Three classifications of associates, namely, high, medium, and low, based upon cultural frequencies of occurrence in the free associational norms, were used in Exp. I for testing the hypothesis. Experiment II employed low-frequency associates and associates having a cultural frequency of 1. A forced-choice test was developed which involved presenting each S with the nonsense syllable stimulus followed by five words, one of which was an associate of the learned word. The S was instructed to select the one word which he felt was most related to the nonsense syllable. The selection by the S of the given associate was assumed to demonstrate the prior establishment of a connection between the associate and the nonsense syllable. The results strongly supported the first hypothesis, but failed to support the second hypothesis as no significant functional relationship was found to exist between the choices of associates in the testing situation and their cultural frequencies of occurrence as responses to the meaningful words.

## REFERENCES

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychol. Rep.*, 1958, **4**, 39–44. (a)

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Verbal generalization: A theoretical rationale and an experimental technique. *Off. Naval Res. tech. Rep.*, 1958, No. 23. (b)

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, **4**, 703–713.

COHEN, B. H. An evaluation of three associational rationales of verbal generalization. Unpublished doctoral dissertation, University of Connecticut, 1958.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

HALL, J. F. Learning as a function of word frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

LUH, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**(3, Whole No. 142).

MORGAN, R. L., & UNDERWOOD, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–603.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WHITMARSH, G. A., & BOUSFIELD, W. A. Use of free associational norms for the prediction of generalization of salivary conditioning to verbal stimuli. *Psychol. Rep.*, 1961, **8**, 91–95.

YAVUZ, H. S., & BOUSFIELD, W. A. Recall of connotative meaning. *Psychol. Rep.*, 1959, **5**, 319–320.

# REVERSAL AND NONREVERSAL SHIFTS WITHIN AND BETWEEN DIMENSIONS IN CONCEPT FORMATION

I. DAVID ISAACS AND CARL P. DUNCAN

*Northwestern University*

In a number of studies of concept learning in human adults (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Mayzner, 1956), Ss were first reinforced for different responses to two stimuli varying on some dimension (e.g., circle vs. square, form dimension) while the stimuli simultaneously varied on one or more momentarily irrelevant dimensions (e.g., color). After mastery of this task (hereafter, the training task), some Ss were shifted to a transfer task in which each of the two stimuli that had been reinforced in training was now paired with the opposite response (reversal shift). Thus, in transfer, reversal Ss had to learn two re-paired S-R associations. Other Ss were shifted to a transfer task provided by reinforcing the stimuli (e.g., red vs. blue) on a previously irrelevant dimension (nonreversal shift to a different dimension). All of the studies using human adults as Ss (those cited above, the only ones of concern here) consistently found that nonreversal shift to a different dimension provided a more difficult transfer task, in terms of trials to learn, than reversal shift.

This finding has been used to support a mediation theory of the way human adults learn and transfer in such concept tasks (for details, see, e.g., Goss, 1961; Kendler & D'Amato, 1955). However, a mediation theory also predicts, according to Kendler and D'Amato, that the reversal condition should yield positive, not negative, transfer in comparison to a control group that learns only the transfer task. Since it is usually found that re-pairing of S-R associations produces negative transfer (e.g., Porter & Duncan, 1953), it is important to determine if this prediction can be confirmed. Of the three studies that used a control group, one (Kendler & D'Amato, 1955) did find that the reversal group learned the transfer task more quickly than the control group; one (Buss, 1953) found the control learned faster than the reversal group; and one (Harrow & Friedman, 1958) found no difference. This disagreement among the studies is probably unimportant because, it is suggested here, none of the studies actually used an appropriate control group. In all cases the control group learned only the transfer task; no attempt was made to equate control and experimental groups on nonspecific transfer variables (e.g., learning to learn, warm up) which would be developed in the experimental groups by the training task. Since nonspecific transfer factors are likely to have a net positive transfer effect, performance of the control groups in the three cited studies was probably poorer than would have been the case if nonspecific transfer had been controlled. So, the present study is a further comparison of reversal shift (R) and nonreversal shift to a different dimension (NRD) in transfer, along with an attempt to provide a more appropriate transfer control for these groups.

In addition to the usual NRD

condition, it is also possible, as Harrow and Friedman (1958) point out, to provide another kind of nonreversal shift in transfer, viz., nonreversal shift on the same dimension that was relevant in training (NRS). Harrow and Friedman suggest that this NRS condition should also, like the R condition, be easier to learn, in transfer, than NRD. This prediction is also tested in the present study.

## METHOD

*Apparatus.*—The $S$ and $E$, seated on opposite sides of a table, were separated by a vertical plywood panel 29 in. high, 48 in. wide. The side of the panel viewed by $S$ was painted gray and contained a plastic window $2\frac{1}{2}$ in. high, $4\frac{1}{2}$ in. wide, centered in the panel 11 in. above the table. Two lights, one on each side of the window, were used to provide reinforcement. Two push buttons were fixed to the table, one below each light. If $S$ pushed either button, the light above it came on to signal a correct choice, provided that $E$ had previously set a mercury switch on $E$'s side of the table.

On $E$'s side of the panel a deck of stimulus cards was pressed against the window by means of a drawbar and springs. Thus, when $E$ removed the card appearing in the window, the next card was immediately revealed.

*Stimuli.*—For $Ss$ in experimental groups the stimuli varied on two dimensions, form and number (of forms), one or the other of which was relevant at some time during the experiment for all $Ss$. In addition, the stimuli varied in color (all stimuli on any one stimulus card were either red or blue), a dimension that was always irrelevant. All stimuli were drawn with colored pencils on white 3 × 5 in. cards. Cards were inserted in plastic envelopes.

There were four values on the form dimension (circle, square, hexagon, triangle), and four on the number dimension (one, two, three, or four forms on a card). At any one time during the experiment $S$ had to respond to just two of the values, on one of the dimensions, paired against each other, e.g., circle vs. square. Only the following pairs were used: circle vs. square, hexagon vs. triangle, one vs. three forms, two vs. four forms.

The training stimuli for the control group were vertical arrows, colored black, drawn on 3 × 5 in. cards. These control stimuli also varied on two dimensions, each relevant for some $Ss$: direction (up-pointing or down-

TABLE 1

STIMULI AND EXPERIMENTAL DESIGN

| Group | $S$ | Training | | Transfer | |
| --- | --- | --- | --- | --- | --- |
| | | Left | Right | Left | Right |
| R (Reversal to same dimension) | 1 | C1, C3 | S1, S3 | S2, S4 | C2, C4 |
| | 2 | C2, C4 | S2, S4 | S1, S3 | C1, C3 |
| | 3 | H1, H3 | T1, T3 | T2, T4 | H2, H4 |
| | 4 | H2, H4 | T2, T4 | T1, T3 | H1, H3 |
| NRD (Nonreversal to different dimension) | 1 | 1S, 1C | 3S, 3C | Same as for Group R | |
| | 2 | 2S, 2C | 4S, 4C | | |
| | 3 | 1H, 1T | 3H, 3T | | |
| | 4 | 2H, 2T | 4H, 4T | | |
| NRS (Nonreversal to same dimension) | 1 | H2, H4 | T2, T4 | Same as for Group R | |
| | 2 | H1, H3 | T1, T3 | | |
| | 3 | C2, C4 | S2, S4 | | |
| | 4 | C1, C3 | S1, S3 | | |
| Control | 1 | UX, UZ | DX, DZ | Same as for Group R | |
| | 2 | XU, XD | ZU, ZD | | |
| | 3 | UX, UZ | DX, DZ | | |
| | 4 | XU, XD | ZU, ZD | | |

Note.—C, S, H, T = circle, square, hexagon, triangle; 1, 2, 3, 4 = number dimension, number of forms on a card; U or D = up-pointing or down-pointing arrow; X = short arrow, Z = tall arrow. Symbols in bold face print indicate reinforced stimuli or dimension. Left and right indicate responses.

pointing arrowhead), and height ($\frac{3}{4}$ in. or 2 in.). There was also a dimension that was always irrelevant, width: an arrow was either $\frac{1}{8}$ in. or $\frac{1}{2}$ in. wide. There was only one arrow on each card.

*Conditions.*—The design is shown in Table 1. There were three experimental groups and a control group, all given different training tasks but the same transfer task. As may be seen in Table 1, all four combinations formed by putting together a *pair* of stimuli from one dimension and a *pair* from the other dimension were used, for different *S*s within each group, in both training and transfer tasks. Group R was trained on a form discrimination, either circle vs. square, or hexagon vs. triangle. Group NRD was trained on a number discrimination, either one vs. three, or two vs. four forms. Group NRS was, like Group R, trained on forms but was transferred to *different* forms, whereas Group R was transferred to the same forms reversed. Group C (control) was trained either on direction (up- or down-pointing arrow), or on height (tall or short arrow). All groups were transferred to the *same* form discrimination.

Table 1 also shows that for Group NRD, partial reinforcement on the transfer task was controlled. When this group was shifted to the transfer task, the previously reinforced values on the number dimension were changed to new values so *S* would not receive partial reinforcement (by continuing to respond to the training stimuli) in transfer. A more detailed presentation of this partial reinforcement issue in concepts shifts is given by Harrow and Friedman (1958). Since, for Group NRD, shifting from training to transfer involved changing stimulus values on one dimension while not changing values on the other dimension, it was decided to use this same "degree of change" for all three experimental groups, as may be seen in Table 1.

There is one more important feature of the design shown in Table 1. It can be seen that in both Group R and Group NRS, any particular form discrimination (e.g., in Table 1, C1, C3 vs. S1, S3) required of some one *S* (subject) during training, was also required of some other *S* during transfer. Therefore, when the training task for either Group R or Group NRS is taken as a whole (ignoring the counterbalancing of stimuli for individual *S*s), each of these training tasks provides a measure of difficulty of the transfer task in the absence of nonspecific transfer from a prior task. In other words, performance of Groups R and NRS during training is a measure of how an

inappropriate control group (not corrected for nonspecific transfer) would perform on the transfer task.

*Subjects.*—The *S*s were students in introductory psychology courses, assigned to groups in turn. Each of the three experimental groups (R, NRD, NRS) was assigned 32 *S*s. Two *S*s in Group NRD failed to reach criterion on the training task and were replaced.

It soon became clear that for *S*s in Group C, discrimination of height of arrows was much more difficult than discrimination of direction. Therefore, 48 *S*s were run in Group C, 24 *S*s on the height discrimination (Group Ch), 24 on direction (Group Cd). Ten *S*s failed to learn the height discrimination and were replaced.

*Procedure.*—For half the *S*s in each group, the stimuli reinforced by pressing the left or the right button are shown under the Left or Right columns in Table 1. For the remaining *S*s this were reversed; stimuli appearing under Left were reinforced for pressing the right button, etc.

With stimuli varying on two dimensions, plus an always irrelevant dimension, a deck of 8 cards was necessary to represent all possible combinations for either experimental or control *S*s. Two such decks were prepared, so 16 cards in all were available. Four different orders of these 16 cards were used. This permitted presentation of each of the four combinations of possibly relevant stimuli, number and form (or height and direction of arrow in Group C), equally often as the first card shown an *S*. Then, for each *S*, the second card presented revealed the other values of number and form (or height and direction) that had not appeared on the first card. Thus, in the first two cards presented *S* saw all possibly relevant stimulus values and dimensions that were to appear in the particular task on which he was working. Four different random orders were used to determine the order of presentation of the remaining 14 cards, with the restriction that the 8 different stimulus cards be shown before any card was repeated.

The instructions to *S* essentially told him to press the left or right button for each card appearing in the window, and that if he were correct, the light above the button would come on.

For both training and transfer tasks, *S* was required to reach a criterion of six successive correct responses. If *S* had not met criterion after three presentations of the pack of 16 cards (six presentations of the 8 different stimulus cards), *S* was dropped as a nonsolver.

The $S$ was allowed to proceed at his own pace; on the average, about 7 sec. elapsed between presentation of successive cards. There was no interruption between training and transfer tasks.

## RESULTS

*Training.*—The left portion of Table 2 summarizes performance on the training task as measured by number of trials to the criterion of six successive correct responses. The six criterion trials are not included in the data. Although 32 $S$s were taken to criterion on the training task in Groups R, NRD, and NRS, 1 $S$ in Group NRS and 1 $S$ in NRD failed to reach criterion on the transfer task. These 2 $S$s were eliminated, and 1 $S$ in Group R, with median performance in training, was also eliminated to reduce $N$ to 31 in each of the three experimental groups.

There was no significant difference among the three experimental groups (top three lines in Table 2) on the training task ($F < 1$). Hartley's test indicated that the group variances were homogeneous ($F_{max} = 1.87$, $df = 3/30$).

It was noted earlier that it was necessary to run separate subgroups in Group C because of differential difficulty of the dimensions of training stimuli. The difference between the mean (see Table 2) of the subgroup that discriminated direction (Group Cd) and the subgroup that discriminated height (Group Ch) was highly significant ($t = 4.06$).

When Group Ch was included with the three experimental groups in analysis of variance of training means, $F$ was 3.22 ($P < .05$, $df = 3/113$). By $t$ test, the mean for Group Ch differed significantly from the means of each of the three experimental groups at the 5% level or less. Analysis of variance of Group Cd and the experimental groups yielded $F < 1$.

## TABLE 2

### MEAN TRIALS TO CRITERION IN TRAINING

| Group | $N$ | Training | | Transfer | |
|---|---|---|---|---|---|
| | | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| R | 31 | 7.03 | 1.70 | 7.52 | 1.72 |
| NRD | 31 | 8.81 | 1.98 | 13.77 | 2.10 |
| NRS | 31 | 7.00 | 1.45 | 3.03 | .93 |
| Cd | 24 | 4.67 | 1.63 | 2.67 | .49 |
| Ch | 24 | 13.83 | 1.63 | 7.42 | 1.76 |

Hereafter, Group Cd will be considered the more appropriate control group.

*Transfer.*—Mean trials to criterion on the transfer task are shown in Table 2. Again, the means do not include the six criterion trials. Since the variances of the three experimental groups were heterogeneous ($F_{max} = 5.03$, $P < .01$), and since the distributions were also positively skewed, the scores were transformed to log ($X + 1$). This eliminated the heterogeneity of variance and produced approximately normal distributions. Analysis of variance of the transformed scores of the experimental groups gave $F = 19.4$ ($P < .001$, $df = 2/90$). By $t$ test, Group R differed significantly from Group NRD ($t = 3.14$), and from Group NRS ($t = 3.32$). Groups NRD and NRS also differed significantly ($t = 6.46$). The fact that Cond. R was easier than Cond. NRD is in agreement with all previous studies that have made this comparison. The new finding is that Cond. NRS was easiest of all.

Analysis of variance of transformed scores of experimental groups and Group Cd yielded $F = 16.6$ ($P < .001$, $df = 3/113$). By $t$ test, Group Cd differed significantly from Group R ($t = 2.42$) and from Group NRD ($t = 5.35$) but not from Group NRS ($t < 1$).

TABLE 3
Mean Error Ratios

| Group | N | Training | | Transfer | |
|---|---|---|---|---|---|
| | | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| R | 31 | .52 | .013 | .70 | .020 |
| NRD | 31 | .53 | .009 | .49 | .007 |
| NRS | 31 | .43 | .013 | .42 | .029 |
| Cd | 24 | .50 | .029 | .44 | .022 |
| Ch | 24 | .50 | .004 | .66 | .020 |

*Errors.*—The number of trials on which $S$ pressed the wrong button (errors), divided by the number of trials to criterion, was computed for each $S$. These error ratios for both training and transfer are summarized in Table 3. There were no significant differences among groups in training. Analysis of variance of the transfer data for experimental groups and Group Cd yielded $F = 4.80$ ($P < .01$, $df = 3/113$). By $t$ test, Group R differed significantly from Group NRD ($t = 2.53$), from Group NRS ($t = 3.39$), and from Group Cd ($t = 2.92$). Other comparisons were not significant.

## Discussion

The data show that Group R, operating under a negative transfer paradigm, did in fact show significant negative transfer when compared to a control group in which nonspecific transfer was controlled. Group R also showed the highest error ratio in transfer, another index of intertask interference.

The need to control for nonspecific transfer in studies of this kind is indicated by the powerful effects such transfer had in the present study. Recall that for a group of $S$s as a whole, the training task for both Groups R and NRS was identical to the transfer task for all groups; therefore, performance of these groups in training yields a measure of difficulty of the transfer task for $S$s not provided with training for nonspecific transfer. This measure was essentially the same for both Groups R and NRS (7.03 and

7.00 mean trials to criterion in training, respectively). Nonspecific transfer was presumably controlled in Group Cd, and this group required a mean of only 2.67 trials (transfer mean) to learn the same task. It seems likely that in the studies of Buss (1953), Harrow and Friedman (1958), and Kendler and D'Amato (1955), the reversal groups would have shown negative transfer had the control groups been trained so as to minimize nonspecific transfer.

Most of the data on reversal shifts in concept learning in human adults has been interpreted in terms of "mediating mechanisms" or "implicit cues" (Goss, 1961; Harrow & Friedman, 1958; Kendler & D'Amato, 1955). The interpretation of the present data, which follows, avoids this particular theoretical language. Instead, the interpretation is based largely on a single, and presumably fairly simple, assumption.

Assume that $S$s reinforced on a particular dimension and extinguished on all other dimensions during training, tend to continue to respond, initially, to the reinforced training dimension during transfer. If so, then Group NRS (trained on forms, transferred to new forms) would have responded primarily to the two new forms on the transfer task. Since the two new forms would have had roughly equal probabilities of association with the two responses, the transfer task would essentially reduce to a simple two-choice discrimination for these $S$s. Group NRS should, and did, learn the transfer task very rapidly.

According to the same assumption, Group R (trained on forms, transferred to the same forms re-paired with the responses) should also have continued to respond to stimuli on the form dimension early in transfer. But because the forms available to these $S$s had been differentially reinforced in training, and were re-paired in transfer, initial probability of association between the forms and responses would not be equal. Thus, although Group R was also faced, it is assumed, with only a two-choice discrimination in transfer, the training associations had to be extinguished

before the transfer task could be learned. Group R should, and did, transfer more slowly than Group NRS, and should make many errors. And as has been shown, Group R should and did transfer more slowly than an appropriate control group.

Still following the basic assumption, Group NRD (trained on number, transferred to forms) would continue to respond to the number dimension during transfer. Since no number stimuli, new or old, were consistently reinforced during transfer, the task for these Ss became quite difficult. First, responses to stimuli on the number dimension had to be extinguished. There now remained two dimensions, form and color, from which to choose; since both these dimensions had been extinguished during training, there was no basis for choosing between them. So Group NRD next had to discover that it was forms, not colors, that was being reinforced during transfer. Finally, these Ss had to discover which form went with which response. It seems clear that the total number of alternatives from which to choose was greater for Group NRD than for any other group (Goss, 1961, has come to the same conclusion), and Group NRD showed the poorest performance of all in transfer. Viewed this way, it is not surprising that Group NRD should be inferior to Group Cd and Group NRS. But in this and in all previous studies that have made the comparison, Group NRD also learned the transfer task more slowly than even the negative transfer group (R). This finding simply shows that having to deal with several stimulus alternatives that have previously been subjected to differential reinforcement and extinction is more difficult than having to deal basically only with two associations, involving basically only two associations, a difference in task difficulty that would seem to have little theoretical import.

## SUMMARY

In a study of human concept formation, two experimental groups were trained on a two-choice form discrimination, with number and color stimuli irrelevant. For one group

(reversal shift) the transfer task consisted of re-pairing the training stimuli with the responses; for the other group (nonreversal shift to the same dimension), two new forms were used as transfer stimuli. A third experimental group (nonreversal to a different dimension) was trained on number stimuli and transferred to forms. A control group was trained on stimuli differing from any of those used for experimental groups, then transferred to forms. The same two-choice form discrimination, with number and color irrelevant, was used as the transfer task for all groups.

The results showed three significantly different levels of performance (in terms of trials to learn) on the transfer task. In order of best to poorest performance, the levels were: (a) nonreversal to same dimension, and control; these groups did not differ, (b) reversal shift, and (c) nonreversal to different dimension. As compared to the control, the reversal group showed significant negative transfer. It was suggested that performance of all groups could largely be accounted for by a combination of two factors: nonspecific transfer, and a specific tendency to continue to respond in transfer to the dimension of stimuli reinforced in training.

## REFERENCES

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

Goss, A. E. Verbal mediating responses and concept formation. *Psychol., Rev.*, 1961, **68**, 248–274.

Harrow, M., & Friedman, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

Kendler, H. H., & Mayzner, M. S., Jr. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

Porter, L., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

# EFFECTS OF SECONDARY REINFORCEMENT SCHEDULES IN EXTINCTION ON CHILDREN'S RESPONDING [1]

N. A. MYERS AND J. L. MYERS

*University of Massachusetts*

Strong secondary reinforcement effects have not been consistently demonstrated. Nor is there agreement regarding the appropriate explanatory concepts. In particular, doubt has been cast upon the explanation of $S^r$ (secondary reinforcement) in terms of "discrimination" between conditioning and extinction trials (Bitterman, Fedderson, & Tyler, 1953).

Support for the discrimination hypothesis comes from a study by Melching (1954). He presented two groups of rats with 50% neutral stimulus (buzzer) in training and found no difference in extinction responding between the group given no buzz in extinction and the group given 100% buzz in extinction. A study by Myers (1960) presents negative evidence for the discrimination hypothesis. She trained children, using tokens as potential secondary reinforcers, and found that of the two groups trained with 50% token, the group receiving 100% token during extinction made significantly more responses than the group receiving no tokens during extinction.

Resolution of the differences in the results of Myers and Melching is difficult without further data. The studies differed in the species of $S$ and in the type of neutral stimulus. The present study was designed to provide a further test of the "discrimination" hypothesis with children as $S$s (as in the Myers' study) and the buzzer as reinforcer (as in Melching's study). Furthermore, a low rate of presentation of both the primary and neutral stimuli has been used during training, in accord with recent data

on the effectiveness of such schedules in establishing secondary reinforcers (Fox & King, 1961; Zimmerman, 1957, 1959). An even more important reason for using such schedules is that the difference between the training rate and 100% buzzer in extinction should be clearly greater than the difference in training rate and 0% in extinction, yielding a better test of the "discrimination" hypothesis than either the Myers or the Melching study.

## METHOD

*Apparatus.*—The apparatus employed was a portable box designed to attract the interest of preschool children. On the front was painted a clown face, having red jewel-light eyes, a push-button nose, and a slot-tray mouth. M & M coated chocolate candy was dispensed through a tube to the mouth of the clown, while a $\frac{1}{2}$-sec. buzz was heard from the interior of the box. The $E$ had access to and operated two silent knife switches which allowed administration of the predetermined reinforcement. The number of responses was recorded on an electric magnetic counter mounted on the back of the box, out of $S$'s sight. The $E$ recorded the number of responses made during each successive minute of extinction.

*Subjects.*—The $S$s were 75 boys and girls between the ages of 4 yr., 7 mo. and 5 yr., 11 mo., attending kindergartens in Northampton, Massachusetts.[2]

*Procedure.*—Each $S$ was asked in the classroom if he would like to play a clown game, and if he acquiesced he was led to a small testing room. The clown was placed

## TABLE 1

DESIGN OF EXPERIMENT AND MEAN NUMBER
OF EXTINCTION RESPONSES
FOR EACH GROUP

| Group | Conditioning | | Extinction | Mean No. of Responses |
|-------|------|------|------|------|
| | Candy (%) | Buzzer (%) | Buzzer (%) | |
| $E_1$ | 20 | 20 | 100 | 102.73 |
| $E_2$ | 20 | 20 | 20 | 55.80 |
| $E_3$ | 20 | 20 | 0 | 36.40 |
| $C_1$ | 20 | 0 | 0 | 72.80 |
| $C_2$ | 20 | 0 | 100 | 66.60 |

on a small table and $S$ was instructed to sit at the small chair in front of it. No written instructions were read to $S$, but $E$ standardized the verbal instructions as much as possible. Attention was called to the clown face, especially the nose. The $S$s were told that "something happens when you press his nose. Let's see what happens." The $E$ pressed the clown's nose and thereby received a buzz and an M & M. The $S$ was encouraged to try it also and was given one rewarded preliminary trial. He was then told he could stay and play the game "as long as you want." The $E$ then sat down behind the table, facing the open back of the box and $S$.

Each $S$ was reinforced according to a predetermined 20% reinforcement schedule which delivered a total of 15 M & M candies. Immediately following the fifteenth reinforcement, the extinction period commenced; no candy reinforcement was administered, and each $S$ was run until he stopped and indicated a desire to return to the classroom or until 5 min. had elapsed, at which time $E$ terminated the session. One last candy was offered at the end of the extinction period.

*Design.*—Eight boys and 7 girls were assigned randomly to each of four groups. Fifteen more children were assigned to a second control group, run after the others. Three E groups received a ½-sec. buzz every time a candy was received during training (20% reinforcement with M & M and buzz). They differed only with respect to extinction treatment: one group (100% buzz) received the buzz for every button press in extinction; one group (20% buzz) heard the buzz on approximately every fifth response, as in training; the third group (0% buzz) *never* heard the buzzer in extinction. A control group ($C_1$) never received the buzzer either during training or extinction; they received 20% reinforcement with M & M candy alone

during training, and no reinforcement during extinction. A second control group ($C_2$) also received 20% reinforcement with M & M candy *alone* during training, but received the buzz for *every* button press in extinction. The design is presented in Table 1, along with the mean number of extinction responses for each group.

## RESULTS

The mean numbers of responses, for successive minutes of extinction, for the five groups are presented in Fig. 1. An analysis of variance was performed on these data and yielded a significant difference between groups ($F = 5.98$, $df = 4/70$, $P < .001$). There was a significant decrease in responding for all groups over time ($F = 60.49$, $df = 4/280$, $P < .001$), but the Groups × Time interaction was not significant.

Duncan's multiple range test was applied to compare the groups with one another. All differences were significant at the .01 level except those between $E_2$ and $C_1$ (where $P < .05$), between $E_2$ and $C_2$, and between $C_1$ and $C_2$.

## DISCUSSION

A simple discrimination explanation of $S^r$ effects (Melching, 1954) would predict greatest number of extinction responses from the 20% buzz group in this study, since the schedule of $S^r$ presentation during conditioning and extinction is identical, and therefore, the extinction period is less discriminable from the
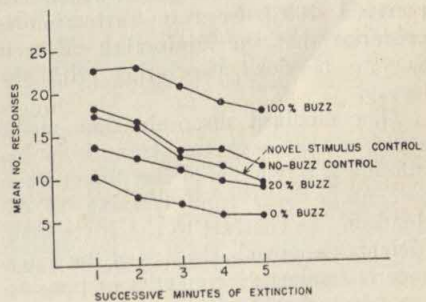


FIG. 1. Mean number of responses for successive minutes of extinction.

former conditioning period than for any other group. Zimmerman (1957, 1959) also would predict greatest response strength for the 20% buzz group, arguing that any $S^r$ value of the buzzer accrued during conditioning would be dissipated more slowly by more occasional presentation during extinction. However, the results quite clearly refute these hypotheses: the 100% buzz group made almost twice as many extinction responses as the 20% buzz group. And, when the 20% buzz group was compared with the primary control group, it was seen that the buzz presented 20% of the time did not operate to increase response strength above primary extinction level. Furthermore, the simple discrimination-generalization model would predict a higher level of extinction responding for the 0% buzz group than for the 100% buzz group, in this experiment, since the change from 20% buzz to 0% buzz is not as great as the change from 20% to 100% buzz, therefore not as discriminable, and conditioned responses should be generalized more easily. Again, the results do not support this prediction: the 100% buzz group made almost three times as many extinction responses as the 0% buzz group.

It appears that some notion of a supplementary reinforcing role of the buzzer stimulus, as suggested by Myers (1958) and Myers (1960) is needed to account for the significantly greater number of responses made with 100% buzz presentation in extinction. It may be noted that the significant difference between the 100% buzz group and the novel-stimulus control group which also received 100% buzz in extinction is evidence that the reinforcing effect is due to previous association with the candy.

This modified discrimination model assumes that response strength in extinction is a function of the difference in percentage buzz from training to extinction. In contrast to the Bitterman-Melching approach, the sign of the difference is retained; increments in percentage buzz should yield more responses than no change, which in turn should

result in more responses than decrements. This prediction is clearly borne out in the present study. The only incompatible finding is the significant difference between Groups $E_2$ (20% buzz in conditioning and extinction) and $C_1$ (0% buzz in conditioning and extinction). The theory would predict no difference, as would also the Bitterman-Melching approach. However, it should be noted that this difference was of considerably less statistical significance than those differences predicted by the theory.

## SUMMARY

Kindergarten children were trained in a free operant situation with candy as a reward. A group receiving 20% buzzer presentations in training, and shifted to 100% buzzer in extinction responded significantly more than similarly trained groups shifted to 20% buzzer and 0% buzzer in extinction. This 100% buzzer group also performed better than a group which was similarly extinguished but which had not experienced the buzzer in training. It was concluded that a secondary reward effect was demonstrated.

## REFERENCES

BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, 66, 456–464.

FOX, R. E., & KING, R. A. The effects of reinforcement scheduling on the strength of a secondary reinforcer. *J. comp. physiol. Psychol.*, 1961, 54, 266–269.

MELCHING, W. H. The acquired reward value of an intermittently presented neutral stimulus. *J. comp. physiol. Psychol.*, 1954, 47, 370–374.

MYERS, J. L. Secondary reinforcement: A review of recent experimentation. *Psychol. Bull.*, 1958, 55, 284–301.

MYERS, N. A. Extinction following partial and continuous primary and secondary reinforcement. *J. exp. Psychol.*, 1960, 60, 172–179.

ZIMMERMAN, D. W. Durable secondary reinforcement: Method and theory. *Psychol. Rev.*, 1957, 64, 373–383.

ZIMMERMAN, D. W. Sustained performance in rats based on secondary reinforcement. *J. comp. physiol. Psychol.*, 1959, 52, 353–358.

# SIMULTANEOUS INDUCTION OF MULTIPLE ANCHOR EFFECTS IN THE JUDGMENT OF FORM [1]

## EDWARD D. TURNER AND WILLIAM BEVAN

### Kansas State University

The traditional approach in psychophysics has been to hold constant all properties of the stimuli to be judged except one and to plot responses as a function of this variable. Meanwhile, perhaps the most obvious characteristic of judgmental situations outside the laboratory is that stimuli to be judged vary among themselves on a number of dimensions. A recent solution to the problem of multidimensionality has been to have stimuli judged for similarity and to express these relationships as distances in a Cartesian space (Torgerson, 1958). An alternative experimental strategy, when the stimulus dimensions can be identified, consists of limiting these dimensions to some small number greater than one, and allowing them to vary with reference to each other in certain prescribed ways. This not only allows for an assessment of the psychophysical relationships involved but may provide some information on the processes of judgment.

The present experiment employs the method of single stimuli. It differs from the usual application in several ways: the stimuli differ with respect to three different physical dimensions; variation on each dimension is independent of variation on the other two; and three judgments, one for each dimension, are made following the presentation of each stimulus.

The purpose of the experiment was to determine whether or not anchoring effects typically obtainable for stimuli varying on a single dimension (Woodworth & Schlosberg, 1954) could also be obtained for multidimensionally varying stimuli.[2]

## METHOD

*Subjects.*—The $S$s were 30 female undergraduates. They were divided randomly into three groups of 10 each. Group A received anchor stimuli which deviated from series stimuli in size and shape but were of a medium lightness. Group B received anchors which deviated in color and shape but which were of a medium size. Group C received anchors which were deviant in color and size but of an intermediate shape. All $S$s judged all stimuli, including anchors, on all three dimensions.

*Stimuli.*—The series stimuli consisted of gray rectangular shapes, each mounted on heavy white (Crescent No. 100) illustration board, 31.5 × 22.5 cm., for presentation in a Gerbrands tachistoscope. The series members differed from each other such that there were 4 each of 4 different shapes, 4 sizes and 4 degrees of lightness. In order to keep the stimulus series to a manageable length, 16 combinations of shape, size, and color (lightness) were selected from the 64 possible. This was done by arranging the 16 possible size-shape combinations in a 4 × 4 matrix and then superimposing on this, in Latin square fashion, the 4 degrees of lightness such that each color appeared in each column and each row only once. The colors were Color-Vu grays No. 7, 9, 10, and 11. Their Munsell equivalents as well as the other physical properties of the stimuli are presented in Table 1.

## TABLE 1

### Physical Properties of the 16 Series Stimuli

| Size (Approx. Area in Cm.²) | Shape (Length × Width in Cm. and Length × Width Ratio) | Lightness (Munsell Number) |
|---|---|---|
| 14 | 3.75 × 3.75 (1:1) | 6.5 |
| | 3.60 × 3.90 (1:1.08) | 5.5 |
| | 3.45 × 4.05 (1:1.21) | 5.0 |
| | 3.30 × 4.70 (1:1.27) | 4.5 |
| 25 | 5.00 × 5.00 (1:1) | 4.5 |
| | 4.80 × 5.20 (1:1.08) | 6.5 |
| | 4.60 × 5.40 (1:1.21) | 5.5 |
| | 4.40 × 5.60 (1:1.27) | 5.0 |
| 38 | 6.25 × 6.25 (1:1) | 5.0 |
| | 6.00 × 6.50 (1:1.08) | 4.5 |
| | 5.75 × 6.75 (1:1.21) | 6.5 |
| | 5.50 × 7.00 (1:1.27) | 5.5 |
| 56 | 7.50 × 7.50 (1:1) | 5.5 |
| | 7.20 × 7.80 (1:1.08) | 5.0 |
| | 6.90 × 8.10 (1:1.21) | 4.5 |
| | 6.60 × 8.40 (1:1.27) | 6.5 |

Two similar stimuli were used as anchors for each group. These represented extreme values on two dimensions and an intermediate value on the third (control) dimension. Group A received the size-shape anchors. These were two relatively large rectangles, 96 cm.² in area, with a length by width ratio of 1:1.50. (Their dimensions were 8 × 12 cm.) They had Munsell values of 5 and 5.5 and thus were intermediate grays. Group A, therefore, provided anchor data on size and shape and control data for color. Group B received the color-shape anchors. These were two black 1:1.50 rectangles of intermediate (4 × 6 cm. and 5.0 × 7.5 cm.) size. Group B thus provided anchor data for the color and shape dimensions and control data for size. Group C received the color-size anchors, two large black rectangles, 9.6 × 10.4 cm. and 9.2 × 10.8 cm. They were intermediate in shape (length to width ratios of 1:108 and 1:21, respectively). Group C provided anchor data on color and size and control data on shape.

Assuming the stimuli designated as anchors to be effective, it was expected that the size judgments for the anchored groups would be reliably smaller than those of the control, the shape judgments would shift toward greater squareness, and the color judgments toward greater lightness.

Each S made a total of 72 judgments for each dimension: each of the 16 series members was presented 3 times and each of the two anchors 12 times. The order of presentation on the 72 trials was random.

*Procedure.*—The Ss were tested individually. Presentations of stimuli were at intervals of 10 sec. for durations of .5 sec. The psychophysical method was the rating scale version of the absolute method. Ratings were required on all three dimensions for each stimulus presentation. Thirteen categories were available for each judgment: The shape categories varied from 0 (perfectly square) to 12 (extremely nonsquare). The size categories were −6 (extremely small) through 0 (neither large nor small) to +6 (extremely large). The color categories varied from −6 (very dark gray) through 0 (neutral gray) to +6 (very light gray). The Ss were also encouraged to use additional categories at either or both ends of any scale when they regarded this to be necessary to the expression of their judgments. The anchors were in no way identified by E as special stimuli. Each S recorded her judgments upon a mimeographed data sheet provided for her. Median judgments were computed for each S's judgments of each individual stimulus. Means of these medians were then used as cell entries in all analyses performed upon the data.

## RESULTS AND DISCUSSION

Figure 1 summarizes the data for each dimension separately. The anchor data consists of the average judgments of the several series stimuli made by two groups; the control data derives from the judgments of the third group. Table 2 presents summaries of analyses of variance used to evaluate these data. Three separate analyses were performed, one on the data of each dimension. In the interest of simplicity of presentation, the within-Ss sources (Between Stimuli, Stimuli × Ss, etc.) have been collapsed, so that the summaries indicate differences between groups of Ss, each of whom is represented by one average judgment per dimension. Similarly, the Stimuli × Groups interaction is not identified. Meanwhile, the between-groups source has been partitioned into predicted differences (an-
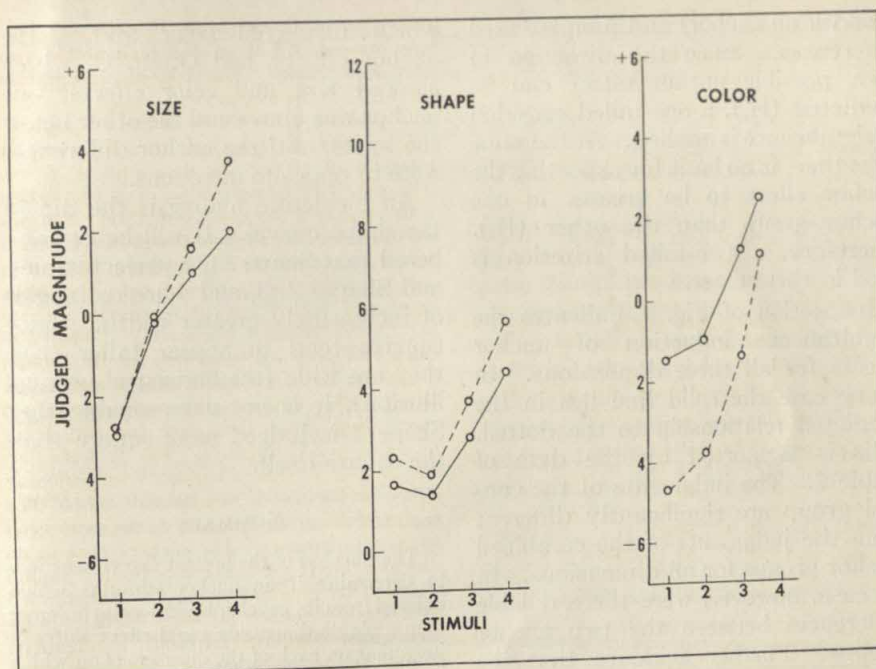
FIG. 1. Average size, shape, and color anchor effects for the several multidimensional groups. (The solid line represents the anchor data, the dotted the control data. The anchor curve for size derives from the data of Groups A and C, its control from Group B. The shape anchor curve represents data of Groups A and B; its control is Group C. The color data are obtained from Groups B and C; its control is Group A.)

TABLE 2

SUMMARIES OF ANALYSES OF VARIANCE PERFORMED UPON THE JUDGMENTS FOR EACH
OF THE THREE DIMENSIONS ON WHICH THE SERIES STIMULI VARIED

| Dimension | Source | df | MS | F | Hypothesis Tested |
|---|---|---|---|---|---|
| Size | Between groups Anchor (Groups A, C) vs. No Anchor (Group B) | 1 | 150.4 | 3.20** | $H_a$ |
| | Group A vs. Group C | 1 | 31.2 | .66* | $H_o$ |
| | Pooled between Ss | 27 | 47.02 | | |
| Shape | Between groups Anchor (Groups A, B) vs. No Anchor (Group C) | 1 | 283.8 | 4.08** | $H_a$ |
| | Group A vs. Group B | 1 | 277.5 | 3.99* | $H_o$ |
| | Pooled between Ss | 27 | 69.6 | | |
| Color | Between groups Anchor (Groups B, C) vs. No Anchor (Group A) | 1 | 2597.0 | 31.90*** | $H_a$ |
| | Group B vs. Group C | 1 | 171.2 | 2.10* | $H_o$ |
| | Pooled between Ss | 27 | 81.41 | | |

\* $P > .05$; two-tailed.
\*\* $P < .05$; one-tailed.
\*\*\* $P < .001$; one-tailed.

chor vs. no anchor) and nonpredicted differences. Since the direction of each possible anchor effect can be predicted ($H_a$), a one-tailed criterion of significance is applied. At the same time there is no basis for expecting the anchor effect to be greater in one anchor group than the other ($H_o$). Therefore, a two-tailed criterion is used in these cases.

Inspection of Fig. 1 indicates the simultaneous induction of anchor effects for all three dimensions. In every case the solid line lies in the predicted relationship to the dotted. This is supported by the data of Table 2. The judgments of the control group are significantly different from the judgments of the combined anchor groups for all dimensions. In no case, however, were there reliable differences between the two anchor groups. Further evidence that the differences between groups are anchor effects is indicated by the difference in slope between each pair of curves. When the anchor is above the series, as in the case of size and shape, the curves should be most widely separated at their upper end; when it is below, the separation should be greatest at the lower end. The data of Fig. 1 are in line with this expectation. Finally, it is interesting to note that simultaneous anchor effects may be either in the same or opposite directions. In the case of Group A, which displayed size and shape anchor effects, both anchors were above the series and the judgmental shifts were downward. However, in Group B,

which displayed color and shape anchor effects, and Group C, which showed size and color effects, one anchor was above and the other below the series, and the anchor differences were in opposite directions.

An incidental finding is the dip in the shape curves. It will be remembered that Shape 1 is a perfect square and Shapes 2, 3, and 4 are rectangles of increasingly greater width. Since squares tend to appear taller than they are wide (the horizontal-vertical illusion), it is not unreasonable that Shape 2 is judged more square than the square itself.

## SUMMARY

The purpose of the present experiment was to determine if an anchor stimulus which differed from its psychophysical series on more than a single dimension could effect shifts in judgment on each of the dimensions on which it differed from the series. Accordingly, $S$s were asked to judge a series of rectangular figures which varied in shape, size, and lightness. Anchor stimuli which represented marked deviations from the series values on two but not on the third dimension were included in the order of presentation. Three groups of $S$s were used so that all combinations of two dimensions were anchored with the third available for control data. Analyses of variance performed on the data for each of the three dimensions indicated that multiple anchoring had occurred.

## REFERENCES

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.
WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

# DISCRIMINATION AND MEDIATED GENERALIZATION IN PROBABILITY LEARNING

## JULIET POPPER SHAFFER [1]

*University of Kansas*

When responses learned to a given stimulus occur in the presence of other stimuli which are not physically similar to the original, we have an example of secondary or mediated stimulus generalization. The generalization in such cases appears to be based on previous experiences of the organism being studied. Some of the best evidence for mediated generalization comes from studies of what has been called semantic generalization, in which a response conditioned to a word generalizes to other words similar in meaning to the original. Reviews of the experimental literature on semantic generalization may be found in Cofer and Foley (1942) and Osgood (1953).

Behavior theorists have attempted to account for these phenomena by assuming that $S$s make implicit responses preceding the overt response and that these implicit responses produce stimuli which partly determine the overt response to the presented stimulus. For example, in the case of generalization from one word to another word similar in meaning, it has been assumed that there are learned mediating responses to words which represent their meanings. The more nearly synonymous two words are, the greater is the similarity of these responses, i.e., the greater is the physical similarity between the patterns of stimulation produced by the

mediating responses. A response learned to a word will also be learned to the stimuli produced by the mediating response representing the meaning of that word. Therefore, by physical similarity, the response will generalize to the stimuli produced by mediating responses to other words similar in meaning to the original, generating a semantic gradient of generalization (Osgood, 1953).

This paper reports the results of an experiment designed to test a model for mediated generalization, developed within the framework of statistical learning theory. The model specifies the assumed mediation process more precisely than has usually been the case, and yields quantitative predictions of the effects of mediation in a specific experimental situation. A brief theoretical review will be given here; for a full account see Popper (1959).

The mediation model is based on a model for discrimination learning developed by Burke and Estes (1957). Their model applies to discrimination problems which consist of a series of trials. Each trial is initiated by a stimulus to which $S$ responds, and is terminated by a reinforcing event. A stimulus is conceptualized as a set of elements available for sampling by $S$, with each element conditioned to one and only one of the response alternatives in the situation. Each available element has a probability $\theta$ of being sampled on a particular trial. The probability of each response is equal to the proportion of sampled elements conditioned to that response.

When a reinforcing event terminates a trial, all elements in the sample become conditioned to the response corresponding to that event.

The mediation model is an extension of the Burke and Estes approach to problems in which mediating responses are assumed to be occurring and influencing the final overt response. Specifically, it is assumed that mediating responses in the presence of a stimulus produce cues which can be represented as additional elements in the set corresponding to that stimulus. These elements are therefore available for sampling when that stimulus is present, and if their conditioning status is known, their effect on the overt response can be predicted.

In this experiment, $S$s were given two successive probabilistic discrimination problems. Their performance on a third problem was predicted on the assumption that it would be affected in a specified way by mediating responses, resulting from the training on the two initial problems.

## METHOD

The first two probabilistic discrimination problems will be designated Discrimination a and Discrimination b, respectively. On each trial of Discrimination a, one of two stimuli, a green light or a white light, appeared, followed by one of two reinforcing events, the letter X or the letter O. Immediately after the light appeared, $S$ was to respond by saying either X or O, to indicate which outcome he expected on that trial. The reinforcing events were probabilistically related to the stimuli, i.e., the probability of X or O on each trial depended only on the stimulus initiating the trial.

The $S$s were then trained on Discrimination b, which was another two-stimulus, two-response problem. The stimuli were X and O, and the reinforcing events were two nonsense syllables, MAF and KUV. Discriminations a and b were related in that the reinforcing events (and responses) of Discrimination a were the same as the stimuli of Discrimination b. Interspersed among the trials on Dis-

crimination b were a few trials on which the stimulus was a green light or a white light, as in Discrimination a, but the $S$ was required to respond with MAF or KUV, as in Discrimination b. No reinforcing event occurred on those trials, which will be referred to as test trials.

Finally, a third problem, Discrimination c, was given in which the stimuli were the green and white lights, the reinforcing events were MAF and KUV, and MAF and KUV each had probability .50 of occurring, regardless of which stimulus initiated the trial.

The stimuli, responses, and reinforcing events in Discrimination a will be denoted by $T_1$ and $T_2$, $A_1$ and $A_2$, and $E_1$ and $E_2$, respectively, where reinforcing event $E_i$ means reinforcement of response $A_i$. Similarly, the stimuli, responses, and reinforcing events in Discrimination b will be denoted by $T_3$ and $T_4$, $A_3$ and $A_4$, and $E_3$ and $E_4$.

*Subjects.*—The $S$s were 96 Indiana University students taking the first semester of introductory psychology. They were assigned randomly to experimental groups, and tested individually.

*Apparatus.*—A vertical black wooden board, 30 in. high and 36 in. wide, was supported on a table 30 in. high. A diffusing screen made of a double layer of sanded Plexiglas, 4 in. high and 21 in. wide, was mounted on the board. Three inches below the center of the screen was a window of one-way mirrored glass, 2 in. in diameter, which became transparent only when lighted from behind. Another window of the same kind was below the first, with 3 in. between the centers of the two windows. A door on the back of the apparatus permitted the insertion of cards immediately behind the windows.

Two 6.3-v. pilot light assemblies were mounted 12 in. apart behind the Plexiglas screen. Colored jewel caps covered the lights so that from a frontal view the left light was green and the right light was white. Two 6.3-v., .15-amp. incandescent bulbs were mounted behind each window, one on each side. A cam-operated timer controlled the time intervals during which the appropriate lights came on.

A 5 × 8 in. index card was behind the windows on each trial. Some of the cards used had either X or O typed in pica capitals so that it would appear in the center of the upper window when illuminated, and either MAF or KUV, typed in pica capitals, so that it would appear in the center of the lower window when illuminated. Other cards were blank in either the upper or lower position.

*Procedure and experimental design.*—Each $S$ sat 3 ft. in front of the table which supported the stimulus panel. The room was dark, except for a 100-w. bulb shining on the front of the windows.

Instructions were given for Discrimination a, presenting it as a prediction experiment and emphasizing the importance of trying to make as many correct choices as possible. After 4 practice trials, $S$s were given 140 trials on Discrimination a, 70 $T_1$ and 70 $T_2$ trials.

Immediately following this phase, $S$s were given instructions for the remainder of the experiment: Discrimination b, with test trials interspersed, and Discrimination c. They were told that any one of four events could begin a trial: the green light or the white light, as before, or X or O. On every trial, they were to guess whether MAF or KUV would follow. They were told also that on some of the trials a blank card would follow their guesses, indicating that they were not being informed of the correct answer for those trials. An additional 98 trials were given, distributed in the following way. Trials 6 and 17 were unreinforced Discrimination b trials, i.e., the stimuli were $T_3$ and $T_4$, in random order, and no reinforcing event occurred. They were included so that $S$s would have some experience with unreinforced trials prior to the test trials, and they have been omitted in all analyses. Trials 33, 45, 56, and 66 were test trials, with stimuli $T_1$ and $T_2$ and no reinforcing events. For half the $S$s in each experimental subgroup, the stimuli appeared in the order $T_1$-$T_2$-$T_2$-$T_1$, and for the rest in the order $T_2$-$T_1$-$T_1$-$T_2$. The remainder of the trials up to Trial 66 were reinforced Discrimination b trials, 30 $T_3$ and 30 $T_4$ trials. Trials 67–98 were Discrimination c trials, 16 $T_1$ and 16 $T_2$ trials.

On each trial of the experiment, the stimulus appeared for 2 sec., followed immediately by the reinforcing event (or a blank white background) for 2 sec., and there was a 6-sec. intertrial interval. The complete experimental session lasted about 45 min.

The probability of reinforcing event $E_j$ following stimulus $T_i$ will be designated $\pi_{ij}$. The $S$s were divided into two main experimental groups. On Discrimination a, for Group I, $\pi_{11}$ was equal to .90 and $\pi_{21}$ was equal to .10, and for Group II, $\pi_{11}$ was equal to 1.00 and $\pi_{21}$ was equal to .50. Except for the different $\pi$ values in Discrimination a, both groups were treated identically; for both, the Discrimination b values were $\pi_{33}$ equal to 1.00 and $\pi_{43}$ equal to .00. The sequences of trials on Discrimination a

were randomized with the restriction that within each successive block of 20 trials, each combination of stimulus and reinforcing event was presented a number of times exactly equal to its expected number, considering the reinforcement probabilities. The sequences of trials on Discrimination b were randomized with the same restriction within each successive block of 10 trials. On Discrimination c, the randomization was restricted in the same manner over the total set of 32 trials. On all problems, a different randomization was used for each $S$. The design was counterbalanced by having eight subgroups, with different identifications of the stimuli and reinforcing events, within each main experimental group, making a total of 16 subgroups, 6 $S$s in each.

## RESULTS

*Discriminations a and b.*—The proportion of $A_j$ responses on $T_i$ trials, within a given block of trials, will be designated $P(A_j | T_i)$. The changes in $P(A_1 | T_1)$ and $P(A_1 | T_2)$ for both groups, over 20-trial blocks on Discrimination a and 10-trial blocks on Discrimination b, are illustrated in Fig. 1 and 2. According to the Burke and Estes discrimination model, the final mean probabilities of response $A_1$ given stimulus $T_1$ and response $A_1$ given stimulus $T_2$ should both be between .10 and .90 for Group I, and above .50 for Group II. However, the final $P(A_1 | T_2)$ for Group II, .44, is significantly below .50 ($t = 2.08$, $P < .05$). In Group I, both final
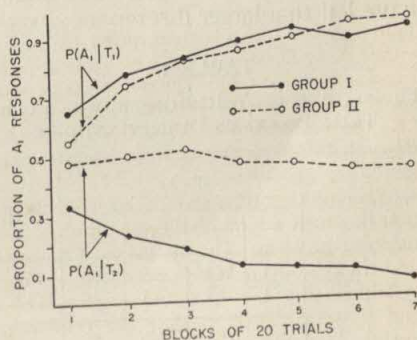


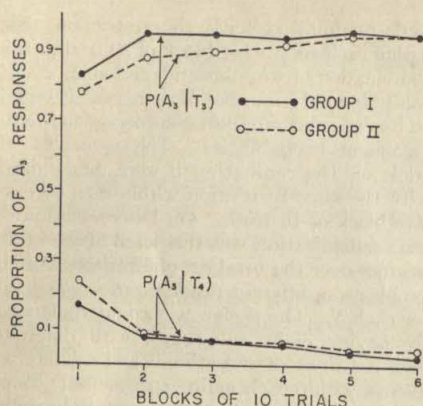FIG. 1. Mean $P(A_1 | T_1)$ and $P(A_1 | T_2)$ over 20-trial blocks for Groups I and II.

FIG. 2. Mean $P(A_3|T_2)$ and $P(A_3|T_4)$ over 10-trial blocks for Groups I and II.

proportions are outside the theoretical limits—$P(A_1|T_1) = .91$, $P(A_1|T_2) = .06$. Since $P(A_1|T_1)$ and $P(A_2|T_2)$, i.e., $1 - P(A_1|T_2)$, are measures obtained under identical experimental conditions in Group I, they were combined in order to get an overall test of the deviation from theoretical bounds in that group. As the distribution of scores is highly skewed, there is no really adequate test for the statistical significance of the deviation from .90. The obtained deviation, .03, is 1.82 times its standard error. Furthermore, a $t$ test of the difference between the proportions on Blocks 6 and 7 indicates a significant increase ($t = 3.21$, $P < .01$), suggesting that continued trials might have led to a larger discrepancy.

*Test trials.*—Two $T_1$ test trials and two $T_2$ test trials were given in order to investigate the dependence among successive unreinforced responses. A preliminary study had shown that a series of unreinforced test trials did not give independent estimates of response probability, since most $S$s adopted a consistent pattern, always making one response to one stimulus and the other response to the other stimulus. Chi square tests were used to investigate response dependence on these trials.[2] The responses on the first $T_1$ test trial and the first $T_2$ test trial did not deviate significantly from independence, while responses on the second test trial with each stimulus were significantly dependent in the direction suggested by the preliminary study. Therefore, only the first $T_1$ test trial and the first $T_2$ test trial for each $S$ were used in testing the predictions derived from the model.

The observed proportions on these test trials were: For Group I, $P(A_1|T_1) = .73$ and $P(A_1|T_2) = .27$; for Group II, $P(A_1|T_1) = .58$ and $P(A_1|T_2) = .44$. The difference between the two proportions was significant for Group I ($\chi^2 = 12.97$, $P < .001$), but not for Group II ($\chi^2 = 1.16$, $P > .10$).

*Discrimination c.*—The 16 $T_1$ trials and the 16 $T_2$ trials on Discrimination c were each divided into four 4-trial blocks, and the proportion of $A_3$ responses in each block was computed for Groups I and II. The results are given in Table 1. It had been expected that, as training progressed on Discrimination c, $P(A_3|T_1)$ and

TABLE 1

PROPORTIONS OF $A_3$ RESPONSES OVER FOUR-TRIAL BLOCKS ON DISCRIMINATION c

| Group | Trials | Blocks | | | | All Trials |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | |
| I | $T_1$ | .64 | .64 | .55 | .51 | .58 |
| | $T_2$ | .34 | .47 | .45 | .44 | .42 |
| II | $T_1$ | .60 | .55 | .62 | .55 | .58 |
| | $T_2$ | .43 | .45 | .53 | .45 | .46 |

[2] The model does not imply strict independence of responses, but the expected degree of dependence cannot easily be determined, and in any case would be so small that the hypotheses of strict independence provide a very close approximation to the predictions which could be derived.

$P(A_3|T_2)$ would change towards .50. To investigate changes, the difference between the proportion of $A_1$ responses on the first 8 trials and the last 8 trials was computed for each type of trial and each group. The significance of the differences was evaluated with $t$ tests, and only the difference for Group I on $T_1$ trials was significant ($t = 2.94, P < .01$). Since so little change occurred over the 32 trials on Discrimination c, the proportions over all trials were used in further analyses.

An analysis of variance was carried out, using the proportions of $A_3$ responses over the trials of Discrimination c, to determine the effects of group, subgroup, type of trial ($T_1$ or $T_2$), and the interactions among these. No effects even approached significance except type of trial ($F = 14.76, P < .001$). Individual $t$ tests indicated that the difference between $P(A_3|T_1)$ and $P(A_3|T_2)$ was significant beyond the .02 level for each of the two experimental groups.

## DISCUSSION

In specifically applying the mediation model to the experiment reported here, the assumed situation on the test trials will be discussed first. On a test trial, a stimulus from Discrimination a was presented, but $S$ was required to make one of the two responses learned in Discrimination b. It is assumed that, on presentation of $T_1$ or $T_2$, $S$ responded implicitly with $A_1$ or $A_2$, the responses conditioned to these stimuli in Discrimination a. The probability of each was assumed to be equal to the probability of the same overt response, given that stimulus, at the end of training on Discrimination a. It is assumed that the implicit response $A_1$ produced stimulus elements which were a subset of the elements associated with the presence of $T_3$ in Discrimination b, and that the same relationship held for $A_2$ and $T_4$.

The probability with which these elements were conditioned to $A_3$ or $A_4$ was determined, therefore, by the training on Discrimination b. The predicted response probabilities on the test trials are therefore a function of the training on both Discriminations a and b.

In Discrimination c, it is assumed that the initial probabilities of responses $A_3$ and $A_4$ were equal to their probabilities on the test trials, and that the probabilities would have gradually approached .50 as training progressed. No predictions about the rate of change can be derived from the model in its present form.

For this experiment, the model implies that $P(A_3|T_1)$ should be greater than $P(A_3|T_2)$ for both groups. This is a result of the reinforcement probabilities on the discrimination problems. In the presence of $T_1$, an implicit $A_1$ response should be more probable than an implicit $A_2$ response in both groups, because of the Discrimination a training. Therefore, with high probability, a subset of the elements associated with stimulus $T_3$ should become available for sampling on those trials, and those elements have a very high probability of being conditioned to Response $A_3$, due to the Discrimination b training. As a consequence, $A_3$ should be the more frequent response on the $T_1$ test trial, and Discrimination c trials. On the other hand, using the same reasoning, $A_3$ should be the less frequent response on the $T_2$ test trial, and Discrimination c trials.

This prediction was confirmed on both the test trials and Discrimination c trials, with only the difference on the test trials for Group II failing to reach statistical significance. This result indicates that a mediation process was occurring during the trials, since the experiment was designed to insure against any possibility that physical similarity could account for the generalization of responses from the stimuli of Discrimination b to those on the test trials and Discrimination c.

The model implies further that there should have been a preponderance of $A_3$ responses on the test trials and Discrimination c trials for Group II. This is

because the average probability of an implicit $A_1$ response across both types of test trials and Discrimination c trials should have been greater than the probability of an implicit $A_2$ response, due to the asymmetrical reinforcement probabilities in Discrimination a. Therefore, the implicit response should have produced a subset of the elements of stimulus $T_3$ more than half the time, making the response $A_3$ more likely than $A_4$. Specifically, the average of $P(A_3|T_1)$ and $P(A_3|T_2)$ should be greater than .50 for Group II. The obtained averages, .51 on the test trials and .52 on the Discrimination c trials, are very close to .50, and their deviations from it do not approach statistical significance.

This failure of the model is reflected in deviations from the quantitative predictions. On the test trials, both proportions for Group I, and $P(A_3|T_2)$ for Group II, were consistent with the specific predictions, while $P(A_3|T_1)$ for Group II was substantially below the predicted value. For derivations and tests relating to the quantitative predictions, see Popper (1959).

The final proportion of $A_1$ responses on $T_2$ trials for Group II was significantly below the minimum value predictable from the Burke and Estes discrimination model. That model implies furthermore that the asymptotic mean proportion of $A_1$ responses for Group II over both $T_1$ and $T_2$ trials should have been .75. The obtained proportion on the last block, .70, was significantly below the predicted proportion ($t = 3.56$, $P < .001$). In an experiment performed by Estes and Burke (1955) testing the model, they used the same $\pi$ values as those for Group II of the present experiment. Their observed mean proportion of $A_1$ responses over the last block of trials was approximately .71 (as estimated from the published curves). Thus, although $P(A_1|T_1)$ and $P(A_1|T_2)$ were considerably different in the Estes and Burke experiment as compared with the present experiment, their mean value was almost the same in the two experiments, and was below the predicted value.

The observed results on Discrimina-

tion a trials deviated from the predictions based on the Burke and Estes model, then, in a specific way: the theoretically more frequent response did not occur as frequently as predicted. The same description would apply to the deviation of the observed results on the test trials and Discrimination c from the predictions derived from the mediation model proposed here. While no explanation for the discrepancy is suggested by the results, the fact that both models err in the same way suggests that some assumptions common to them are inadequate in this experimental context. Since all of the assumptions of the Burke and Estes discrimination learning model are incorporated into the mediation model, modification of the more general assumptions of the discrimination model seems to hold the greatest promise for achieving a more adequate quantitative formulation of the process of mediated generalization.

## SUMMARY

This experiment was designed to test a quantitative model, based on statistical learning theory, for mediated generalization. The $Ss$ were given training on two discrimination problems (a and b). These problems consisted of a series of trials, each trial beginning with the appearance of one of two stimuli, with $Ss$ required to guess on each trial which one of two possible outcomes would follow the presented stimulus. Each outcome had a prearranged probability of following each of the stimuli. Discriminations a and b were related in that the possible outcomes on Discrimination a were the stimuli with which trials began on Discrimination b. The $Ss$ were then given Discrimination c, and their performance on it was predicted on the assumption that it would be affected in a specified manner by mediating responses resulting from the training on Discriminations a and b. Specifically, the trials of Discrimination c began with presentation of one of the stimuli from Discrimination a, with $Ss$ required to guess which of the two outcomes used in Discrimination b would follow. The probabilities of their initial guesses in this case were predicted on the assumption that they would first respond covertly on the basis of the outcomes of Discrimination a, and that their covert response would produce internal stimulation similar to the corresponding stim-

ulus on Discrimination b. Stimuli from the covert responses would therefore mediate generalization of the responses learned in Discrimination b to Discrimination c.

The results indicated a significant effect of the pretraining on the final problem, along the lines predicted from the model. The precise quantitative predictions were only partially confirmed. The discrepancies between observed and predicted results were compared with discrepancies of a similar nature between observed data on discrimination problems and predictions based on a statistical model for discrimination learning.

## REFERENCES

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimina-tion learning. *Psychometrika*, 1957, **22**, 133–145.

COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.

ESTES, W. K., & BURKE, C. J. Application of a statistical model to simple discrimina-tion learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

OSGOOD, C. E. *Method and theory in experi-mental psychology.* New York: Oxford Univer. Press, 1953.

POPPER, J. Mediated generalization. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford Univer. Press, 1959. Pp. 94–108.

# SEMANTIC SATIATION AND PAIRED-ASSOCIATE LEARNING [1]

## R. N. KANUNGO,[2] W. E. LAMBERT, AND S. M. MAUER

### McGill University

The phenomenon of satiation has been described by Smith and Raygor (1956) as "the reduction in the effectiveness of a stimulus with continued exposure." Two different methods, have been used to produce the satiation effect on verbal stimuli. One involves the overt verbal repetition of the stimulus while the other relies on prolonged visual exposure to the stimulus. The verbal satiation effect has also been observed in various ways. For instance, Basette and Warne (1919) reported lapses of the meaning of words following their verbal repetition, and, more recently, Lambert and Jakobovits (1960) reported measurable decrements in the intensity of semantic ratings of continuously repeated words. Using the prolonged visual exposure method, Smith and Raygor (1956) demonstrated that a word loses its familiarity in the sense that associational responses to a stimulus word become uncommon.

The present studies explored the role of the satiation process in paired-associate learning. The main question considered was whether the reduction of the meaning of words has a detrimental effect on subsequent acquisition tasks involving those very words (Exp. I). In view of the role that meaning plays in the response

position of the paired-associate tasks (Cieutat, Stockwell, & Noble, 1958), it was decided to administer the satiation treatment to response elements of S-R pairs. A second experiment (Exp. II) was performed to study the role of interpolated semantic satiation on the recall of responses of the paired associates.

## EXPERIMENT I

### Method

*Subjects.*—The Ss were 30 undergraduate students. None had previously participated in a similar experiment.

*Material and apparatus.*—Using nonsense syllables and words as stimulus and response members, respectively, two lists of paired associates, each containing eight pairs, were prepared. Nonsense syllables were chosen from Hull's list of less than 20% association value (Hilgard, 1958), and the response words were chosen on the basis of their high frequency of usage (Thorndike & Lorge, 1944) and their high connotative meaning (Jenkins, Russell, & Suci, 1958). Each list was printed on a strip of paper in five different random orders in a manner suited to the standard anticipation procedure with a memory drum. The stimulus term alone was presented for 3 sec. and immediately following it the stimulus-response pair was presented for 3 sec. Then followed the next stimulus exposed for 3 sec. and so on. The intertrial interval was 6 sec.

Another eight words were chosen as controls on the same basis as described above for response words, except that each of them was made equal in length to a response word of the second list. These words were used as controls in the sense that they were not to enter into the learning task after they had been given satiation treatment. Care was taken that the control words were neither structurally nor semantically related to the response words of the paired associates which were to be learned.

Three semantic differential scales (Good-

Bad, Active-Passive, Strong-Weak) representing the three major factors of connotative meaning (Osgood, Suci, & Tannenbaum, 1957) were used for measuring the intensity of semantic ratings of words. Each paired-associate response word and control word was printed on a separate 3 × 5 in. index card. Each semantic scale was also printed on a separate card. All cards were placed in a Kardex folder so that E could expose them in a predetermined random order, one at a time, first a word, and then a semantic scale along which S gave his ratings of the immediately preceding word.

*Procedure.*—All 30 Ss were tested individually. Initially, S was presented the first paired-associate list (List I) with standard instructions for the anticipation procedure involving the use of a memory drum. Before the actual presentation of the list, S was made familiar with the anticipation procedure by a single presentation of two practice pairs.

Three consecutive successful anticipations were considered as the learning criterion. On the basis of their learning scores, Groups C (control) and E (experimental), equated for both trials and errors, were formed for the main stage of the experiment. There were 15 Ss in each group.

The main part occurred approximately 1 wk. after each S's initial testing. For each S of Group E, the normal semantic profile was obtained for each of the eight response words of the second paired-associate list (List II). The procedure was the same as that used by Lambert and Jakobovits (1960). Briefly, each word was exposed for 1 sec. and then S was asked to indicate the appropriate semantic placement by pointing to one of the seven positions on the semantic scale. Then, for the satiation treatment, each of the response words was again exposed for 1 sec., and S was asked to repeat the word aloud continuously for 15 sec., at a rate of 3–4 repetitions per sec. Immediately after the repetition, E exposed a semantic scale and S made his rating for the word. This procedure was repeated three times for each of the eight words, one time for each semantic scale. The words and the scales were presented in an order which maximized the separation of reoccurrence of a word and a scale. For Group C, however, the eight control words were used instead of the List II response words. From each S of Group C, first, the normal semantic profile was obtained for each of the control words, and then satiation treatment was administered to these words. Thus the Ss of Group C were given exactly the same

type of treatment as given to Group E, except that the eight words which were rated and satiated were not those to appear as response words in the paired-associate list.

Immediately after the satiation treatment, each S of Groups E and C was presented the second paired-associate list on the memory drum with exactly the same instructions as given for learning List I. The same procedure and learning criterion as described for the initial stage were used again.

*Results*

Both the trial and the error measures for learning of List I make it clear that Groups C and E were in fact equated for the main stage of the experiment. The mean number of trials to reach criterion for Group C was 10.20 (SD = 3.17), and for Group E was 10.07 (SD = 2.46). Likewise, the mean error scores for Groups C and E were 20.00 (SD = 12.58) and 20.07 (SD = 9.86), respectively.

An examination of Table 1 indicates that for Group C, the satiation treatment of the control words led to a significant decrement in their rated meaning. For Group E however, the meaning decrement does not quite reach significance (.05 < P < .10). A $t$ test applied to the mean satiation scores of both groups revealed no reliable differential effect of the satiation treatment on the two groups ($t = .55$). Since Groups C and E do not differ significantly with respect to their satiation scores, the data from both the groups were combined to see if the overall effect of satiation treatment is to reduce the meaning intensity of the words. The combined mean semantic rating scores presented in Table 1 show that the meaning decrement is significant ($P < .01$).

The effect of satiation of response words on the acquisition of the second paired-associate list is shown in Table 2. Group C, given satiation treatment for control words immediately before learning, was significantly su-

TABLE 1

EFFECT OF SATIATION TREATMENT ON THE SEMANTIC PLACEMENT OF WORDS

| Group | N | Before Satiation | | After Satiation | | Change | | |
|---|---|---|---|---|---|---|---|---|
| | | Mean[a] | SD | Mean | SD | Mean | SD_Diff. | t |
| C | 15 | 4.20 | 1.68 | 3.88 | 1.94 | 0.32 | 0.52 | 2.29** |
| E | 15 | 4.68 | 1.77 | 4.47 | 1.70 | 0.21 | 0.41 | 1.94* |
| C + E | 30 | 4.44 | 1.90 | 4.18 | 1.84 | 0.26 | 0.47 | 3.30*** |

[a] Entries are average polarity scores per word over the sum of three semantic scales.
* $.05 < P < .10$.
** $P < .05$.
*** $P < .01$.

perio to Group E with respect to acquisition of the list. In terms of error scores the difference between the groups is significant beyond the .01 level, but in terms of trials to criterion, the difference is not reliable ($.05 < P < .10$).

*Discussion*

Two general conclusions can be drawn from the results of the study. First, in support of the earlier findings of Lambert and Jakobovits (1960), the study shows that the overall effect of the satiation treatment of words is to reduce the intensity of their meaning. The reason for not obtaining a significant satiation effect in Group E, while Group C showed such an effect, is unclear. However, there is one possibility. It will be observed that in Group E, the initial ratings of the response words are higher

TABLE 2

EFFECT OF SATIATION TREATMENT OF RESPONSE WORDS ON THE LEARNING OF PAIRED ASSOCIATES

| Group | N | Trials | | Errors | |
|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD |
| C | 15 | 6.47 | 2.12 | 8.67 | 5.11 |
| E | 15 | 8.20 | 2.45 | 14.47 | 5.35 |
| t | | 1.20* | | 2.94*** | |

* $.05 < P < .10$.
*** $P < .01$.

than the initial ratings of the control words in Group C (see Table 1). Such higher semantic ratings imply greater polarization of judgments on the part of Ss in Group E. According to Osgood, Suci, and Tannenbaum (1957, pp. 155 ff.) polarization of judgments is an index of habit strength. Thus it would be expected that in Group E the "word-meaning" habit is stronger than the similar habit in Group C. Consequently, Group E would show stronger resistance than Group C to any semantic change as a result of satiation treatment.

The second and most interesting finding is that satiation treatment applied to response words has a negative transfer effect on the later learning of a paired-associate list. Lambert and Jakobovits (1960) conceptualized the phenomenon of semantic satiation as "a cognitive form of reactive inhibition" and related it to Osgood's theory of representational mediation processes. Their explanation could account for the superiority of Group C over Group E in paired-associate learning by assuming that reduction in the meaning of response members makes them more difficult to associate. However, the results can also be accounted for in terms of principles of associative learning. When a response member (R) is continuously repeated, the different associations elicited by the word (m components) may gradually extinguish whereas the R-R connection gets strengthened. This could be an instance where experimentally developed frequency of stimulation (n) may lead to

decrease in $m$. Decrease in meaning as a function of satiation treatment, therefore, can be interpreted in terms of increasing $S$'s tendency to connect the word with itself rather than to any of its common associates.

Thus the effect of satiation of response words on subsequent acquisition can be interpreted in terms of transfer from one learning situation to another. For the experimental group, the meaning of the response words decreased possibly because of the formation of an association of the response word with itself which would produce an impairment in the subsequent learning of the paired associates. The situation is analogous to developing R-R connections for the experimental group where all the $m$ components ("hooks" or associations) of R extinguished, and similarly X-X connections for the control group where all the $m$ components of R remain unaffected before S-R learning. Extinction of $m$ components of R before learning for the experimental group would explain the superiority of the control group. The importance of $m$ components in verbal learning is well recognized (Noble, 1952).

More recently Cieutat (1960) in trying to clarify some of the conflicting data concerning the locus of familiarization and its effect on paired-associate learning, noted that, "familiarity only with the response member inhibits learning" (p. 274). It should be observed that his method of familiarization involved continued visual presentation for 60 sec. similar to the prolonged visual exposure method of satiation. To explain his results he argues "that the monotony of continued visual presentation evokes an inhibiting influence" (p. 274).

Another possible interpretation of the present results makes use of response similarity. The prelearning satiation treatment given to the response words reduced their meaning, possibly making them more alike semantically. If so, one would expect to find more intralist response competition for Group E than for Group C. An examination of errors revealed that 61% of all errors for Group E are intralist intrusions in comparison with 67% for Group C, a comparison which rules out this interpretation.

## Experiment II

We were interested in extending this line of reasoning to another aspect of verbal learning. The present study compared the effects of the satiation treatment on stimulus and response members of paired associates when the treatment was presented after the associates had been learned. In this case both stimulus and response members were meaningful words. Use was made of a simple retroactive inhibition design. During the original learning phase, the S-R connections were established, while during the interpolated phase either stimulus (for one group) or response elements (for a second group) were given the satiation treatment, and finally recall of response elements was tested when stimuli were presented.

### Method

*Subjects.*—The $S$s were 52 university students. None had previously participated in an experiment of this type.

*Materials and apparatus.*—Several quite different methodological procedures were employed in Exp. II. Using meaningful words as stimulus and response members, a list of 12 paired associates was prepared. The words were chosen on the basis of their high frequency of usage in print (Thorndike & Lorge, 1944) and their high connotative meaning (Jenkins et al., 1958). Each of the 12 pairs was judged (by 12 students acting as judges) to have little or no immediate association between its stimulus and response members.

Each paired associate was printed on a separate 3 × 5 in. card. Further, each stimulus and response member was printed on a separate card. These cards were placed in a Kardex folder so that E could expose them in a predetermined random order. Each stimulus word was placed immediately before the paired associate to which it corresponded so that E could expose the stimulus-response pair after the exposure of the stimulus word in a reliably constant manner with a minimum of delay.

Three semantic scales were used for semantic ratings. These were: Good-Bad, Active-Passive, Strong-Weak.

*Procedure.*—The study used two test conditions, a "Stimulus condition" and a "Response condition." Each test condition was in the form of a retroactive inhibition paradigm and was divided into three phases.

*Learning phase.*—This phase was identical for both test conditions. Each S was given four trials, a complete trial consisting of the exposure, in a predetermined, random order, of each stimulus member of the paired associates followed by the stimulus-response pair. Each stimulus member and each pair was exposed for 3 sec. and a 10-sec. delay was given between trials.

After four learning trials Ss were assigned to either the Stimulus or Response condition depending on their learning efficiency, equating the two groups on paired-associate learning ability.

*Stimulus condition.*—First, S's normal semantic profiles for all 12 stimulus words were obtained. Each word was presented three times (for 1 sec. each time) for measurement on the three semantic scales. The words and scales were also presented in a predetermined randomized order.

[Each of the 12 stimulus words was placed in one of two categories, Satiation Category (SC) or Nonsatiation Category (NSC). An attempt was made to group one half of the stimulus members of paired associates which had been learned by the fourth learning trial in SC, and the other half in NSC. Cases where odd numbers of associations had been learned were balanced through the total group. Further, one half of the stimulus members of paired associates which had not

been learned by the fourth learning trial were grouped in SC, the other half in NSC.

Each word in SC was exposed for 1 sec. and Ss were asked to repeat the word aloud for 15 sec. at a rate of 2–3 repetitions per sec. Immediately after the continual repetition, Ss rated the word on one of the three semantic scales. Each word in NSC was exposed for 1 sec. and Ss rated it immediately after exposure. After the list had been subjected to this treatment once (each word in SC receiving satiation treatment and being measured on one scale, and each word in NSC merely measured on one scale) all words were then rated in the usual way on the remaining scales. That is, each stimulus word was exposed for 1 sec. and then rated immediately on one of the two remaining scales. Note that the satiation treatment was only given once, before one of the semantic ratings, not before each rating as was the case in Exp. I. Initial and final semantic ratings were subsequently compared.

*Response condition.*—The procedure for this condition was identical to that for the Stimulus condition except that the response rather than the stimulus members were grouped into SC or NSC categories and then given the satiation treatment.

It can be seen from this procedure that words in SC and words in NSC were exposed an equal number of times to Ss. Furthermore, due to the equal division of the words belonging to correctly learned paired associates into SC and NSC in each test condition, a basis was established for comparing the effects of satiation and nonsatiation treatments on the recall of learned paired associates. Likewise, due to the division of the study into two test conditions, a basis was created for comparing

TABLE 3

AVERAGE CHANGE IN POLARITY OF PAIRED-ASSOCIATE MEMBERS
OVER THE SUM OF THREE SCALES: EXP. II

| Condition | First Rating | | Second Rating | | Change | | |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | t |
| Stimulus | | | | | | | |
| Satiated | 4.77 | 1.05 | 4.18 | 1.25 | .59 | .99 | 2.95*** |
| Nonsatiated | 4.89 | 1.15 | 5.09 | 1.16 | .20 | .62 | 1.66 |
| Response | | | | | | | |
| Satiated | 4.97 | 1.32 | 4.46 | 1.26 | .51 | .73 | 3.40*** |
| Nonsatiated | 4.60 | 1.27 | 4.66 | 1.38 | .06 | .61 | .50 |

Note.—Twenty-six Ss took part in each of the test conditions (Stimulus and Response).
*** P < .01.

TABLE 4

EFFECT OF SATIATION TREATMENT OF PAIRED-ASSOCIATE MEMBERS
ON THE RECALL OF PAIRED ASSOCIATES

| Condition | Mean Scores | | Drop in Recall | | SC vs. NSC Words | | |
|---|---|---|---|---|---|---|---|
| | On Trial 4 | After Interpolation | Mean | SD | Mean$_{Diff.}$ | SD$_{Diff.}$ | $t$ |
| Stimulus | | | | | | | |
| SC Words | 3.00 | 1.73 | 1.27 | .86 | .69 | .91 | 3.78**** |
| NSC Words | 2.85 | 2.27 | 0.58 | .84 | | | |
| Response | | | | | | | |
| SC Words | 2.65 | 1.96 | 0.69 | .82 | .12 | .85 | 0.68 |
| NSC Words | 2.65 | 1.84 | 0.81 | .68 | | | |

Note.—Twenty-six $Ss$ took part in each of the conditions.
**** $P < .001$.

the effect of satiation treatment given to stimulus and response words on their recall.

*Recall stage.*—This stage of the study was identical for both test conditions. The $Ss$ were shown each stimulus word for 3 sec. and asked to recall the response word paired with it.

## Results

Table 3 presents the mean change in polarity scores for stimulus and response words, respectively. It can be seen that in both cases the reduction in intensity of meaning as measured by the semantic differential is significant for words given satiation treatment ($P < .01$ for both stimulus words and response words). On the other hand, words not given satiation treatment showed no significant semantic change.

Table 4 presents the mean number of paired associates learned by the fourth trial. In the Stimulus condition, an attempt was made to administer interpolated satiation treatment to half of the stimulus members of these learned paired associates and not to the other half. A similar attempt was also made in the Response condition except that instead of stimulus members, the response members of the learned paired associates received the interpolated treat-

ments. An examination of Table 4 reveals that such an attempt was successful. The mean number of correct responses on the recall trial after the interpolation treatments, also presented in Table 4, reveals how much interference resulted from the satiation or no satiation treatments of the words learned by the fourth trial. In the Stimulus condition, a mean drop of 1.27 in recall of responses of the learned paired associates is noticed when the stimulus members of those paired associates are given satiation treatment. But the mean drop in the response recall of the learned paired associates of which the stimulus members were in NSC is .58. The difference between these means is highly significant ($P < .001$).

It can be seen that the mean drop in recall scores for learned paired associates of which the response members were given satiation treatment was .69 and the mean drop in recall for learned paired associates whose response members were in NSC is .81. The difference between these two means, of course, was not significant.

Some of the paired associates which were originally unavailable to $Ss$ after four learning trials were available at

recall. It is difficult to speculate as to whether these paired associates were at an "oscillation period" of availability (Osgood, 1953, pp. 503–504) or whether they were learned during the fourth trial when the correct response to the stimulus was exposed, or whether they were somehow made available during the interpolated period. Whatever the source of learning, its pattern is consistent with the other results. Of the 30 paired associates unavailable after four trials in the stimulus condition which were subsequently available at recall (a total of 30 paired associates for the group) 19 were ones whose stimulus members were in NSC while only 11 were ones whose stimulus members were in SC. Further, of the 25 paired associates unavailable after four trials in the Response condition which were available at recall, 16 were ones whose responses were in the NSC while 9 were ones whose responses were in the SC. These observations clearly follow the trends established by the results presented in Table 4.

## Discussion

The findings of Exp. II demonstrate that paired-associate connections can be retroactively disrupted if the connotative meanings of their stimulus members are satiated. However, associational bonds are not affected by satiating response members of already learned paired associates. These results could be explained in terms of the associational interpretation of semantic satiation presented earlier in connection with Exp. I. Here it is argued that continual repetition of a word (TABLE, TABLE, TABLE, etc.) would strengthen the tendency for the word TABLE to be made as a response to the stimulus word TABLE. Thus, if the interpolated satiation treatment involves formation of a positive reaction tendency or a word-word habit, then in the present experiment, the stimulus satiation condi-

tion can be considered analogous to the response variation retroaction paradigm, and retroactive interference would be expected (Osgood, 1953, pp. 525 ff.). On the other hand, the response satiation condition is analogous to the stimulus variation retroaction paradigm where retroactive facilitation is expected. The reason why retroactive facilitation could not be obtained in the response satiation condition must depend upon factors other than the formation of the word-word habit per se during the interpolated period. In view of the importance of meaning in the response positions of the paired-associate tasks, it seems logical to presume that the reduction in meaning of the response items during the interpolated period might have counteracted the facilitating effect of the word-word habit. This explanation, however, leads to the theoretical expectation that the retroactive facilitation effect can be obtained after interpolated satiation treatment to the response items if one uses nonsense verbal units as responses. The findings of Exp. II when considered with the findings of Exp. I, make it clear that the effects of reduction of meaning of response items on the formation (as in Exp. I) or on the maintenance (as in Exp. II) of associational bonds are always detrimental.

## SUMMARY

The role of verbal satiation in paired-associate learning was investigated. Two groups of 15 *S*s each were matched on the basis of their learning measures in an initial test using a paired-associate list. In the main test both the groups learned a second paired-associate list. But immediately before learning, Group E (experimental) was given satiation treatment of the response members while Group C (control) was given similar treatment to words which were not response members. Results indicated that (*a*) the satiation treatment of words caused a decrease in their connotative meaning as measured on semantic scales and (*b*) Group E was slower in learning than Group C.

In Exp. II, using a retroactive inhibition paradigm, the effect of satiation treatment of stimulus words on the recall of already learned paired associates was studied. Satia-

tion treatment resulted in significantly more retroactive interference than did the non-satiation control treatment. The interpolated satiation of response words produced no significant effect on later recall. Satiation treatment given to both the stimulus and the response words resulted in a significant reduction in the intensity of their meanings as measured by semantic differential scales.

The results were discussed in terms of an associational interpretation of semantic satiation.

## REFERENCES

BASETTE, M. F., & WARNE, C. J. On the lapse of verbal meaning with repetition. *Amer. J. Psychol.*, 1919, **30**, 415–418.

CIEUTAT, V. J. Differential familiarity with stimulus and response in paired-associate learning. *Percept. mot. Skills*, 1960, **11**, 269–275.

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness ($m$, $m'$) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1958. Pp. 517–567.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

LAMBERT, W. E., & JAKOBOVITS, L. A. Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.*, 1960, **60**, 376–383.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

SMITH, D. E. P., & RAYGOR, A. L. Verbal satiation and personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 323–326.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words*. New York: Teachers College, Columbia University, 1944.

# EFFECTS OF VISUAL AND VERBAL CUES ON LEARNING A MOTOR SKILL [1]

## LAWRENCE KARLIN AND RUDOLF G. MORTIMER [2]

*New York University*

In the training of motor skills additional cues may be supplied that will not be present in the operational situation. Improvement during training produced by such cues has frequently been found not to persist in subsequent tests in which these cues were not present. On the basis of such results, Miller (1953) has distinguished between cues that tell $S$ what to do next, which he labels "action feedback," and cues that tell $S$ what he should have done, which he labels "learning feedback." The same cue may function in varying degree both as action and as learning feedback. According to Miller, cues which function primarily as action feedback do not produce improvement in performance in tests from which they have been removed, and they may even produce a decrement relative to control conditions.

A study by Lincoln (1954) on the effects of different cues on learning to turn a crank at a specified rate is relevant to the above distinction. One group was given verbal information on the amount and direction of the average rate error after each training trial. A second group was given this information plus a continuous visual cue during each training trial which indicated instantaneous rate error. Both groups yielded similar learning curves but in criterion (retention) tests in which only the intrinsic kinesthetic cues remained, the verbal group did significantly better than the verbal-visual group.

These results may mean that the visual cue functioning as action feedback was a useful guide to performance but did not promote learning to use the intrinsic kinesthetic cues. The verbal cue when used alone may have functioned as learning feedback but in combination, the visual cue functioning as action feedback was so much more available that it minimized the use of the verbal cue.

Continuing along these lines a study by Karlin (1960) investigated the effects of visual, auditory, kinesthetic, and verbal error cues, both singly and in a number of combinations, on performance of a task similar to that used by Lincoln (1954). It was found that a combined visual and verbal cue was consistently but not significantly superior to a verbal cue in learning, and equally good in retention.

These results did not agree with those obtained by Lincoln, and suggested that certain differences between the experimental conditions might be important. Thus while the verbal cue used in both studies was the same, the visual cue was continuous in Lincoln's study and discrete in Karlin's study. It is possible that the failure to find similar results in retention was due to the fact that the discrete visual cue was less informative than Lincoln's continuous visual cue. One of the objectives of the present investigation (Karlin & Mortimer, 1961) was to

check this possibility by using a continuous visual cue both alone and in combination with a verbal cue.

In order to gain further knowledge concerning the mode of action and differential effectiveness of the visual and verbal cues, a scoring system was used by which performance could be evaluated in terms of both constant and variable errors. It was felt that this technique would prove valuable in determining the underlying effects of feedback. Lincoln also had scored performance for both constant and variable errors but he did not obtain significant results for the variable errors. Variable errors were measured by the number of times $S$ passed in and out of the tolerance range. This method of measuring variable error is contaminated with constant error factors which Lincoln may have disregarded because he felt that the variable error would be practically important when measured by its effect on a total accuracy score only when the constant error was relatively small. In the present study the apparatus was specifically designed to yield constant and variable error scores which were independent of each other.

## METHOD

*Subjects.*—The $S$s were 45 paid, volunteer, right-handed male college students.

*Apparatus.*—Except for the use of a continuous visual display the apparatus was basically the same as that used by Karlin (1960), where a more detailed description may be found. Essentially, the apparatus consisted of a crank handle 1 in. in diameter and 5 in. long, masked from $S$'s view, which turned on a mainshaft at a diameter of 7 in. Connected to the mainshaft was a Weston tachometer generator whose output was fed into the electronic scoring system and into the display meter.

The scoring system consisted of 15 channels in which high speed counters cumulated the time that $S$ was turning at a rate within the range that defined each channel.

The display consisted of a Triplett voltmeter carrying a translucent 4 × 2 in. scale, illuminated from the rear, and graduated into 50 units with a center zero marking. The meter was mounted in a vertical panel 22 in. in front of $S$. The meter responded to the output of the tachometer generator (which had a linear response function) such that at 99 rpm the meter needle would be at the center of the scale as indicated by the zero mark.

*Procedure.*—The $S$s were seated in front of the crank assembly and grasped the crank with the right hand. The task was to learn to turn the crank at 99 rpm.

The $S$s were randomly assigned to the visual, verbal-visual, and verbal cue conditions, 15 $S$s per condition. Those receiving the visual cue were instructed in the use of the display meter. Those receiving the verbal cue were informed, at the end of a trial, of the amount and direction of the mean rate error, in rpm. The verbal-visual group was given both types of cue. During retention trials the feedback cues were removed.

The $S$s were tested on 2 consecutive days. On Day 1 $S$s received 3 practice trials without feedback, 25 learning trials with feedback, 15 immediate retention trials, and 10 relearning trials with feedback. On Day 2 they received 15 delayed retention trials and 15 relearning trials with feedback. The first session lasted approximately 50 min. and the second session, which took place about 24 hr. later, lasted 30 min.

A buzzer was used to indicate the beginning of a trial and 3 sec. after $S$ began to turn the crank scoring was begun. At the end of a further 15 sec. a Hunter timer broke the scoring circuit and the buzzer was sounded to inform $S$ of the end of a trial. The intertrial interval was 30 sec. The interval between blocks of trials was about 2 min. All $S$s wore headphones to muffle outside noise. Masking noise was provided by a fan.

## RESULTS

Three measures of performance were obtained for each $S$ on each trial as follows: (a) Total time (sec.) that $S$ turned at a rate within the tolerance range of ±13.5 rpm. (b) Constant error (rpm); i.e., the arithmetic mean of the rate errors, which gave the average amount by which $S$ was turning too slow or too fast on each
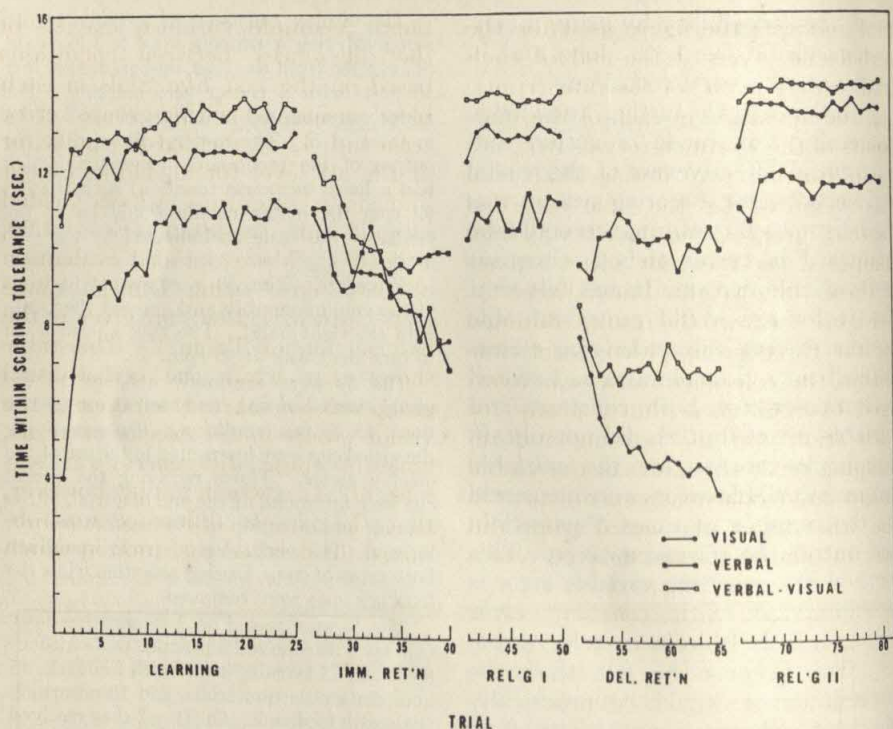
FIG. 1. Mean time within tolerance range by trial and cue ($N = 15$ each cue condition).

TABLE 1

DUNCAN RANGE TESTS OF MEAN DIFFERENCES BETWEEN CUES WITHIN BLOCKS
OF TRIALS FOR DIFFERENT SCORING TECHNIQUES

| Comparison | Learning | | Immediate Retention | | Relearning I | | Delayed Retention | | Relearning II | |
|---|---|---|---|---|---|---|---|---|---|---|
| | D | P | D | P | D | P | D | P | D | P |
| Time Within Scoring Tolerance in Seconds | | | | | | | | | | |
| Verb: Verb-Vis | −2.61 | .01 | 1.87 | ns | −2.83 | .01 | 3.17 | .05 | −2.51 | .01 |
| Verb: Vis | −1.73 | .01 | 1.95 | ns | −1.97 | .01 | 5.77 | .01 | −1.94 | .01 |
| Verb-Vis: Vis | .88 | ns | .08 | ns | .86 | ns | 2.60 | ns | .57 | ns |
| CE in rpm | | | | | | | | | | |
| Verb: Verb-Vis | 2.87 | .01 | −5.68 | .01 | 2.07 | .01 | −7.80 | .05 | 1.47 | .01 |
| Verb: Vis | 1.40 | .05 | −5.61 | .01 | 1.07 | .05 | −11.47 | .01 | .60 | ns |
| Verb-Vis: Vis | −1.47 | .05 | .07 | ns | −1.00 | .05 | −3.67 | ns | −.87 | .01 |
| SD in rpm | | | | | | | | | | |
| Verb: Verb-Vis | 3.33 | .01 | 3.86 | .01 | 4.59 | .01 | 3.00 | .05 | 3.93 | .01 |
| Verb: Vis | 2.27 | .01 | 1.80 | .01 | 3.26 | .01 | 2.60 | .05 | 3.00 | .01 |
| Verb-Vis: Vis | −1.06 | .01 | −2.06 | ns | −1.33 | ns | −.40 | ns | −.93 | ns |

trial and was the figure used for the verbal cue given at the end of each trial. (c) The SD of the rate (rpm); i.e., the deviation of each of the midpoints of the 15 rate intervals from S's mean turning rate during that trial was weighted by the time recorded for that interval, and the SD was then computed as the root mean square of these time-weighted deviations.

Figure 1 shows the results obtained for the three groups when total time within the tolerance range was averaged over all Ss within a group for each trial. In the learning and relearning trials the scores of the verbal group are consistently poorest. On the other hand, the verbal group did best in immediate and delayed reten-

tion. A simple variance analysis of the differences between conditions based on the last five trials in each block using a within-groups error term with 42 df, yielded Fs significant at the .01 level for all blocks except immediate retention, which yielded insignificant results ($F = 1.99$, $P > .05$). More detailed evaluation of these data using Duncan range tests (Edwards, 1960) are given in the first section of Table 1. This table shows that while the verbal-visual group was consistently superior to the visual group in all blocks of trials, none of these differences was significant. It is worth noting, however, that the largest difference was obtained for delayed retention in which
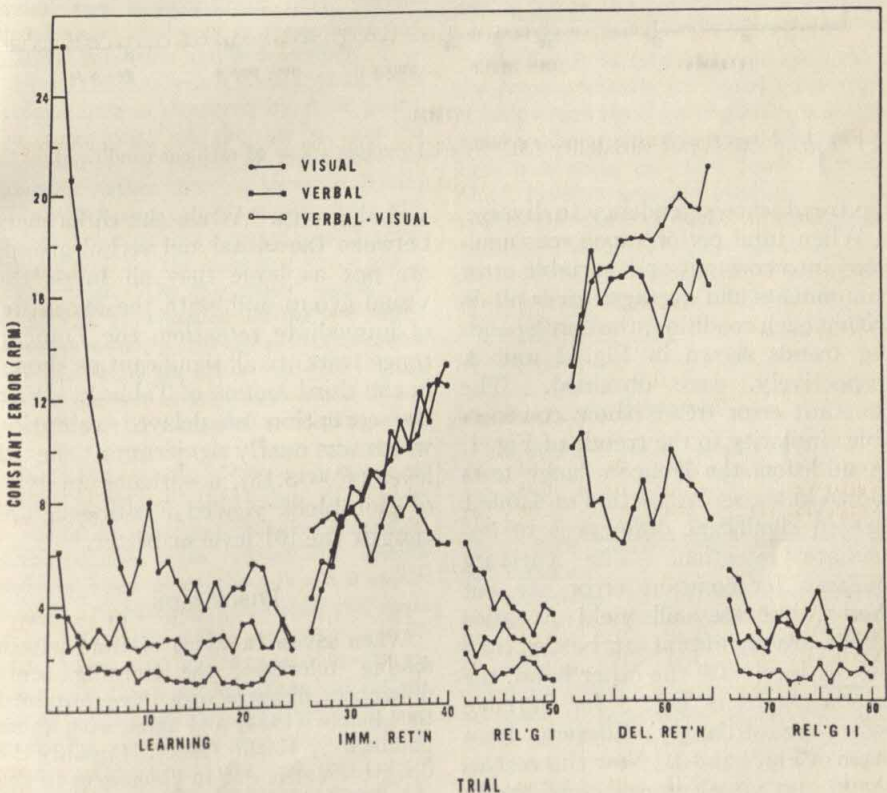


FIG. 2. Absolute mean of constant errors by trial and cue ($N = 15$ each cue condition).
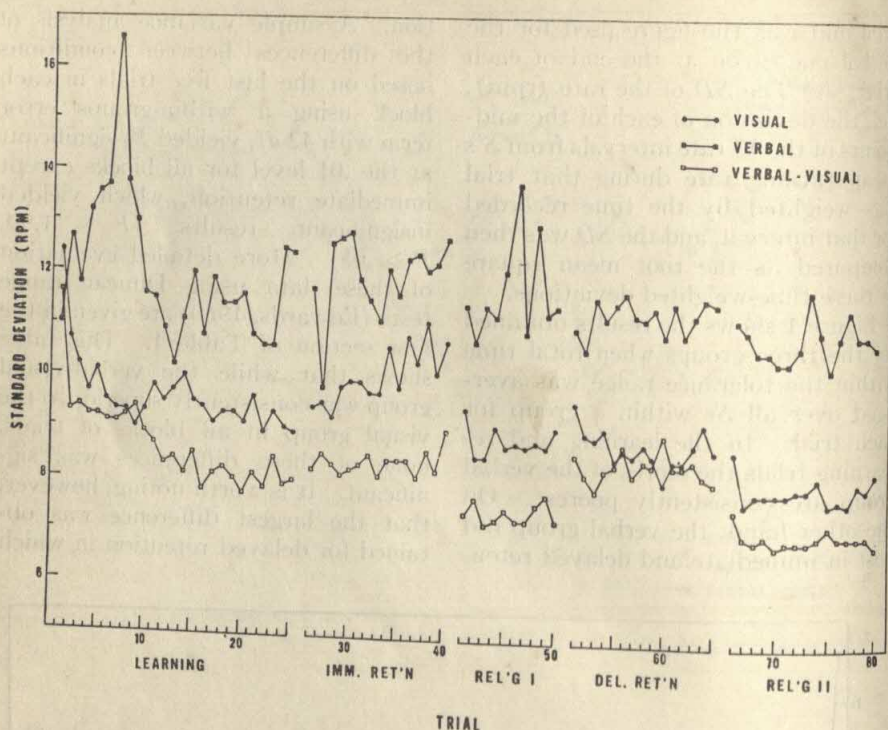
Fig. 3. Mean rate variability (*SD*) by trial and cue (*N* = 15 each cue condition).

the trends show a tendency to diverge.

When total performance was analyzed into constant and variable error components and averaged over all *S*s within each condition, the corresponding trends shown in Fig. 2 and 3, respectively, were obtained. The constant error trends show considerable similarity to the trends of Fig. 1. In addition, the Duncan range tests shown in the second section of Table 1 yielded significant differences in immediate retention. The variance analyses for constant error are not shown, but they all yield *F* ratios which are significant at better than the .01 level. On the other hand, the trends shown in Fig. 3 for variable error are strikingly different from those of Fig. 1 and 2. Now the verbal-visual and visual groups consistently yield lower variable errors than the

verbal group. While the differences between the visual and verbal groups are not as large they all favor the visual group and, with the exception of immediate retention, the Duncan range tests are all significant as shown in the third section of Table 1. With the exception of delayed retention which was nearly significant at the .05 level (*F* = 3.15), a variance analysis of each block yielded *F* ratios significant at the .01 level or better.

## Discussion

When given in terms of time within scoring tolerance, the learning score differences disagree with those obtained by Lincoln (1954) and agree with those obtained by Karlin (1960). Possibly the disagreement is due to differences in the characteristics of the display although the present display differed appreciably

from those used in both of the above studies.

In immediate retention these results agree with Lincoln's findings that the verbal cue was superior to the verbal-visual cue, although in the present study this difference was not significant.

When performance is further analyzed into constant and variable error components, differences among the conditions are more pronounced. Considering the constant error first, the verbal cue is significantly superior to the verbal-visual cue in retention and significantly inferior to this cue in learning. A similar picture is obtained when the verbal cue is compared to the visual cue although the differences in the learning and relearning trials are not so great. Significant differences are also obtained (with one exception) when performance is analyzed in terms of variable error but this time the verbal cue is inferior to both verbal-visual and visual cues in immediate and delayed retention and in learning.

From this analysis it is clear that when performance is measured by total score, the superiority of the verbal cue in retention is a result of its effect on constant rather than on variable errors. This conclusion is reasonable since the verbal cue provided information directly determined by the constant error for a given trial. On the other hand, the visual cue provided an immediate index of performance which did not distinguish between constant and variable errors since it did not average over time. However, the results suggest something that was not apparent in those of earlier experiments, namely, that with "action" cues (see Miller, 1953) like the visual or verbal-visual cues something is learned that reduces rate variability which persists even after the cues are withdrawn. Apparently, the visual cue leads to a relatively stable improvement in smoothness and steadiness of performance. On this point note that the verbal-visual cue is superior both in learning and retention to the visual cue. These results suggest that the two cues may interact to produce greater steadiness during retention than the visual cue alone

but since the results are not statistically significant further work on this question is needed.

It is worth noting that the verbal cue in Lincoln's as well as in the present study did not give variable error information and one may speculate on how a verbal cue which gave an average of the variable errors at the end of a trial would affect performance.

On the whole the results of the present study show that cues which might ordinarily be considered to function as action feedback, or as a "crutch" to guide performance, can make a contribution to retention of a motor skill by way of reducing variable error, although this contribution can be obscured if only the total score is considered. On the other hand the results support Lincoln's finding that the visual cue can be a "hindrance" when combined with the verbal cue as far as the constant error component of the total score is concerned.

Finally, it is important to note that the present results are based on a type of task which involves producing a single steady state. In this sense they may have a bearing on other types of task which involve a single production such as the line-drawing tasks of Thorndike (1932). Further evidence for this conclusion is to be found in a recent study by Baker and Lavery (1960) who used a series of tasks requiring a single end product.

## SUMMARY

The effects of visual, verbal, and combined verbal-visual cues on the learning and retention of a crank-turning task were investigated. The task was to turn the crank at 99 rpm. The Ss were 45 right-handed males, 15 Ss in each condition.

It was found that: (a) Overall superiority in retention tests of task performance measured by time within tolerance was due mainly to reduction of constant errors. (b) The verbal cue was inferior in learning but superior in retention tests when performance was measured by time within tolerance and magnitude of constant error. (c) The verbal cue was inferior to the visual and combined verbal-visual cues during both learning and

retention trials when variable error was measured.

## References

BAKER, C. H., & LAVERY, J. J. Performance during training as a criterion of retention of motor skills. Report, 1960, Defense Research Board of Canada.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rhinehart, 1960.

KARLIN, L. Psychological study of motor skills: Phase I. *USN Train. Dev. Cent. tech. Rep.*, 1960, No. 558–1.

KARLIN, L., & MORTIMER, R. G. Psychological study of motor skills: Phase II. *USN Train. Dev. Cent. tech. Rep.*, 1961, No. 558–2.

LINCOLN, R. S. Learning a rate of movement. *J. exp. Psychol.*, 1954, **47**, 465–470.

MILLER, R. B. Handbook on training and training equipment design. *USAF WADC tech. Rep.*, 1953, No. 53–136.

THORNDIKE, E. L. *Fundamentals of learning.* New York: Teachers College, Columbia University, Bureau of Publications, 1932.

# PREDICTION OF SOME STOCHASTIC EVENTS:
## A REGRET EQUALIZATION MODEL [1]

### MAX S. SCHOEFFLER

*Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey*

The present paper describes a series of experiments which were performed to develop some data on human prediction capabilities, a model for how such prediction capabilities are expressed as behavior, and a series of tests of this model.

The data were collected in a rather simple experimental situation. People were asked to make "bids" on 100 successive numbers that appeared. Either they were instructed that these numbers referred to "make-believe dollars" (MBDs) which they might win, or they were told only to guess as close as possible to the next number. For groups that bid for MBDs, the number of MBDs $S$ received depended on the relationship between his bid and the number that actually occurred (and thus on the adequacy with which he was able to predict the upcoming number).

All of the experiments to be reported here involved this general situation. Two of these experiments provided the intuitive basis for constructing a model of behavior under these circumstances. The remaining experiments were then used to evaluate the adequacy of the model when the assumptions made in constructing the model were specifically tested for generality under conditions to which the model seemed applicable.

One variable that was investigated concerned the specific relationship between the bid that was made and the

number that came up. It was termed the payoff variable and was varied for different groups in order to provide one test of the adequacy of model.

The condition henceforth labeled Guess is the one characterized above as not receiving MBDs. For the Nonpunish condition, the payoff to $S$ was

$$P = \begin{cases} B & \text{if } B \leq N \\ O & \text{if } B > N \end{cases}$$

where $B$ is the bid made by $S$, $N$ is the number that appeared (input), and $P$ is the payoff to $S$ in MBDs. That is, $S$ won as many MBDs as he had bid so long as his bid was less than or equal to the input number. If his bid exceeded the input number, he won nothing. Similarly, for the Punish condition,

$$P = \begin{cases} B & \text{if } B \leq N \\ -B & \text{if } B > N \end{cases}$$

Thus the punish condition differed from the Nonpunish condition only in that a person lost the amount that he bid in case of an overbid, rather than simply receiving nothing.

Make-believe dollars were used because some pilot work indicated that they would function adequately as incentives. In an attempt to retain a linear value scale for the MBDs, only a relatively small range of values was used. To the extent that the value remained linear with number of MBDs, the paradigm permitted a reasonable specification (in this arbitrary unit of measurement) of the payoff matrix relating each response ($B_i$, $i = 1, \ldots, m$) to each experimental outcome ($N_j$, $j = 1, \ldots, k$) in terms of the payoff, $P_{ij}$, (positive or negative) to $S$. For the present study involving predicted numbers as responses, $B_i = i$ and $N_j = j$.

The model that was devised to de-

TABLE 1

INPUT DISTRIBUTIONS FOR THE VARIOUS EXPERIMENTAL GROUPS

| Group | $N$ | $I^a$ | Payoff Function | Trials | | | |
|-------|-----|-------|-----------------|--------|--------|--------|--------|
| | | | | 1–30 | 31–60 | 61–90 | 91–100 |
| G1a | 14 | C | Guess | 8–15 | 8–15 | 8–15 | 8–15 |
| G1b | 16 | | | | | | |
| G2a | 21 | S | Guess | 8–15 | 13–20 | 8–15 | 12 |
| G2b | 20 | | | | | | |
| G3 | 20 | C | Guess | 13–20 | 13–20 | 13–20 | 13–20 |
| N1a | 16 | C | Nonpunish | 8–15 | 8–15 | 8–15 | 8–15 |
| N1b | 24 | | | | | | |
| N2a | 18 | S | Nonpunish | 8–15 | 13–20 | 8–15 | 12 |
| N2b | 20 | | | | | | |
| N3 | 19 | C | Nonpunish | 13–20 | 13–20 | 13–20 | 13–20 |
| P1a | 17 | S | Punish | 8–15 | 13–20 | 8–15 | 12 |
| P1b | 15 | | | | | | |
| P2a | 15 | S | Punish | 13–20 | 18–25 | 13–20 | 17 |
| P2b | 13 | | | | | | |

$^a$ Input distribution Constant (C) or Shifting (S).

scribe behavior in this situation treats "regret" as a central concept. Regret is defined as the difference between the payoff on a trial and the maximum possible. That is, if $P_{ij}$ represents the payoff to $S$ given response $B_i$ and event $N_j$, then the regret experienced by $S$ is $R_{ij} = |P_{ij} - \max_i P_{ij}|$.

Although for the guess condition $P_{ij}$ is not explicitly defined, a concept of regret still appears intuitively meaningful. This regret is considered to be the difference between the predicted number and the actual number. Formally, $R_{ij} = |B_i - N_j|$.

Under this definition, the regret outcomes of a trial must take on positive values. However, since the experiment under consideration deals with ordered outcomes and the responses can likewise be ordered, it is reasonable to consider separately a regret due to overbidding and one due to underbidding. That is to say, since the values of $B_i$ and $N_j$ have been, respectively, identified with the numbers represented by $i$ and $j$, then $R_{ij}$ is due to overbidding if $i > j$, is due to underbidding if $j > i$, and is zero if $i = j$.

The *regret equalization hypothesis* to be proposed as a model, involves the notion that regret due to overbidding and regret due to underbidding result, respectively, in tendencies to lower or raise the response. That is, if for example, $R_{ij}(n)$ is the regret on Trial $n$ and $i > j$, then the response on Trial $n + 1$ tends to be less than $i$; if $i < j$, then the response on $n + 1$ tends to be greater than $i$. If another assumption is made, viz., that the amount of this effect is linear with $R_{ij}$, it seems reasonable to expect behavior to stabilize (if indeed it does stabilize) at a point where the expected value of the regret due to overbidding is equal to the expected value of the regret due to underbidding.

More formally stated, the regret equalization hypothesis predicts an asymptotic bid level $b$ such that $b$ satisfies

$$E(R_{ij}) = \sum_{j=[b]}^{\infty} p_j R_{bj}$$
$$= \sum_{j=0}^{[b-1]} p_j R_{bj} = E(R_{ij})$$

where $[b]$ is the smallest integer $\geq b$ and $p_j$ is the probability that $N_j$ occurs.

It is clear that $b$ is not necessarily an

integer, whereas the $N_j$ are integers. Thus, $b$ cannot represent a constant terminal response level for an individual. Rather it must be an average taken at least over several responses of an individual. Further, the same asymptote is predicted for all individuals facing the same sequence of events. No doubt individual differences do exist, but for the sake of a model with maximal simplicity and intuitive appeal, it was deemed proper to avoid introducing a parameter to deal with individual differences. Rather an attempt will be made to evaluate the predictions using *no* fitted parameters. (This is perhaps an overstatement since the hypothesis was constructed using the data from two of the groups. Thus, it may be argued that the transformation from MBDs to regret in itself constitutes fitting a parameter.)

## METHOD

All experiments used groups of between 13 and 24 college students which constituted classes in freshman, sophomore, or junior psychology courses. The experiments were all conducted in a similar fashion and required that $S$ predict on 100 successive trials the numbers that $E$ wrote on the blackboard on those trials. The instructions to $Ss$ differentiated the three payoff functions that were used.

In the Punish and Nonpunish groups, $Ss$ were instructed that they would get 100 successive opportunities to request between 0 and 30 make-believe dollars (MBDs). On each trial they were to write down the amount of money they were requesting on that trial. After they wrote this number, $E$ wrote a number on the blackboard. For the Punish groups the instructions stated that if the amount requested was less than or equal to the amount that $E$ wrote on the board $S$ would receive the amount requested. If, however, $S$ requested more than $E$ subsequently wrote, $S$ would lose the amount that he bid. For the Nonpunish groups the penalty for overbidding was that $S$ simply did not win any MBDs on that trial and otherwise the rules were the same as for the punish condition. For the Guess groups, $Ss$ were simply told to guess "as close as possible to the number that $E$ would write on the blackboard, with over- and underestimations being equally bad."

These differences in instructions provided to the different groups constituted the payoff variable.

For each bid, each $S$ entered on his data sheet the amount he bid, whether he won or lost on that trial, and a running total of his winnings. (However, $Ss$ in the Guess groups only wrote down the amount bid.) Bids were requested and amounts were written on the board by $E$ at the rate of about one bid every 15 sec. In all, the experiment involved 100 such trials.

The numbers that $E$ wrote on the blackboard are called "input." The numbers were randomly selected from particular distributions. The distributions from which they were drawn differed for the different blocks of trials. For all experiments the distributions were rectangular, the integers included all appearing with equal probability. In Table 1 the distributions are indicated by the highest and lowest integer used. For example, 8–15 represents the integers 8, 9, 10, 11, 12, 13, 14, and 15.

## RESULTS AND DISCUSSION

*Payoff variable: Guess.*—Curves describing performance under the Guess condition are presented in Fig. 1 and 2. Figure 1 presents the means and inter-$S$ $SD$s for Groups G1a, G1b, and



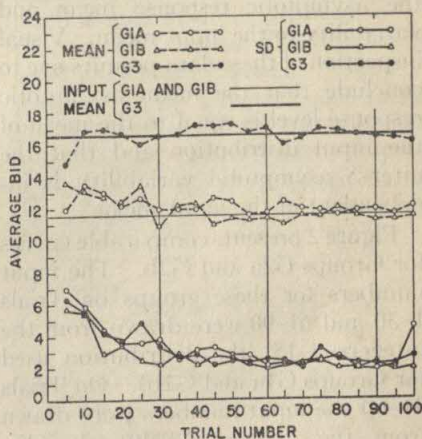FIG. 1. Means and inter-$S$ $SD$s for the Guess condition with a *constant* input distribution, averaged over five-trial blocks. (Also shown are the averages of the input distributions. These coincide with the predicted asymptotic response levels.)

Fig. 2. Means and inter-$S$ $SD$s for the Guess condition with a *shifting* input distribution, averaged over five-trial blocks. (Also shown are the averages of the input distributions. These coincide with the predicted asymptotic response levels.)
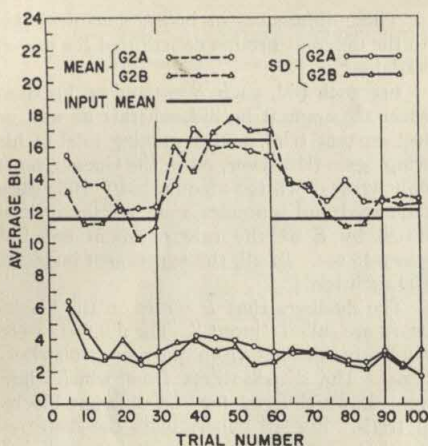
**G3.** As indicated in Table 1, the Groups G1a and G1b differed procedurally from Group G3 only in having their input distributions displaced by five units. Thus, these data serve to indicate the relationship of the asymptotic response mean and variability to the input mean. Visual inspection of these data permits one to conclude that the mean asymptotic response level is equal to the mean of the input distribution and that the inter-$S$ asymptotic variability is independent of the input mean.

Figure 2 presents comparable curves for Groups G2a and G2b. The input numbers for these groups on Trials 1–30 and 61–90 were drawn from the integers 8–15 (the distribution used for Groups G1a and G1b). On Trials 31–60 the input numbers were drawn from the integers 13–20 (the distribution used for Group G3). The input numbers of Trials 91–100 were always 12, but the corresponding data will not be discussed for any of the groups.

The data in Fig. 2 can thus be compared to the data from Fig. 1. They indicate that the groups exposed to shifts in the input distribution approach asymptotes comparable to those attained by groups maintained on the same input distribution throughout. The inter-$S$ variability is a negatively accelerated decreasing function of trials, except that if a shift is introduced, the variability apparently increases temporarily.

Under this (Guess) condition, the regret equalization hypothesis requires that the mean asymptotic response, $b$, satisfy

$$\sum_{j=0}^{[b-1]} (b - N_j)p_j = \sum_{j=[b]}^{\infty} (N_j - b)p_j$$

where $[b]$ is the smallest integer $\geq b$, $N_j$ is an input number and $p_j$ is the probability that $N_j$ is chosen on a given trial. Solving this equation yields $b = \sum_{n=0}^{\infty} N_j p_j = \bar{N}_j$, the input mean.

The mean response curves thus are in line with the hypothesis. However, the symmetry of the situation is such that almost any conceptual model would make a similar prediction. Consequently, the predictions of the model are next examined under the conditions similar to those used here, but with an asymmetric value structure superimposed.

*Payoff variable: Nonpunish.*—Figures 3 and 4 describe the performances of the groups that were given the Nonpunish instructions. The data of Fig. 3 are analogous to those of Fig. 1 and the data of Fig. 4 to those of Fig. 2. Groups N1a and N1b differed from Group N3 in that N3 received an input distribution that was displaced by five units. Groups N2a and N2b received the same input dis-

tribution as Groups N1a and N1b on Trials 1–30 and 61–90, and the same distribution as Group N3 on Trials 31–60.

The mean response data are again characterized by apparently very stable asymptotes when the same input distribution is used throughout (Fig. 3). Again the differences between N1 and N3 in the mean asymptotic response demonstrate the dependence of the response mean on the mean of the input distribution; while the similarity in the inter-$S$ variability between these groups indicates that this measure is independent of the input mean. However, the response asymptotes are no longer at the mean of the respective inputs, but rather are at a new value appreciably below the mean of the input.

In the case of the shifting distribution (Fig. 4), the response asymptotes are again predictable from the data of the groups receiving the same input distribution throughout. Again there is apparently a slight increment in
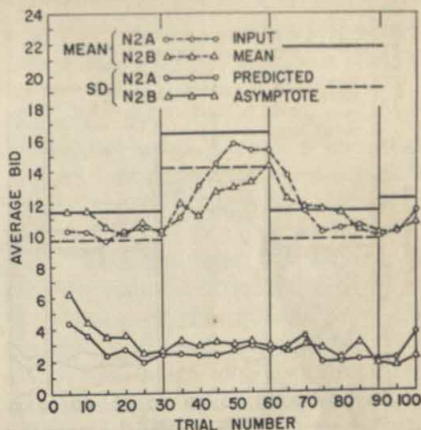


FIG. 4. Means and inter-$S$ $SD$s for the Nonpunish condition with a *shifting* input distribution averaged over five-trial blocks. (Also shown are the averages of the input distributions and the predicted asymptote for each group. The input distribution is the same as that of Fig. 2 for the Guess condition.)

response variability associated with a shift in the input distribution and a negatively accelerated monotonic decrease in variability when the input distribution is not shifted.

According to the regret equalization hypothesis, the asymptotic response mean should again be predictable. If the response is less than the corresponding input number, there is assumed to be a regret equal to the difference between the two. In case of an overbid, since $S$ receives nothing, his regret is equal to the input. Thus, his asymptotic bid level $b$ should satisfy

$$\sum_{j=0}^{[b-1]} N_j p_j = \sum_{j=[b]}^{\infty} (N_j - b) p_j$$

where the right side of the equation constitutes the expected regret due to underbidding and the left side constitutes the expected regret due to overbidding.

Under the Nonpunish condition for the input 8–15, the equation is
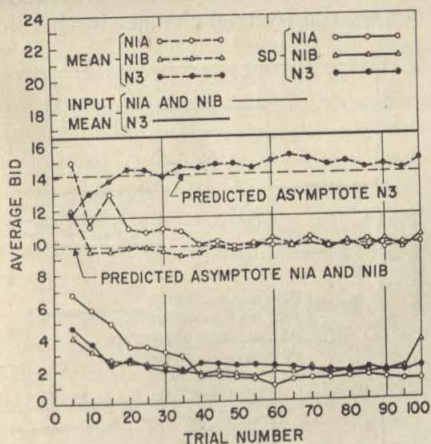


FIG. 3. Means and inter-$S$ $SD$s for the Nonpunish condition with a *constant* input distribution averaged over five-trial blocks. (Also shown are the averages of the input distributions and the predicted asymptote for each group.)

FIG. 5. Means and inter-$S$ SDs for the Punish condition with a *shifting* input distribution averaged over five-trial blocks. (Also shown are the averages of the inp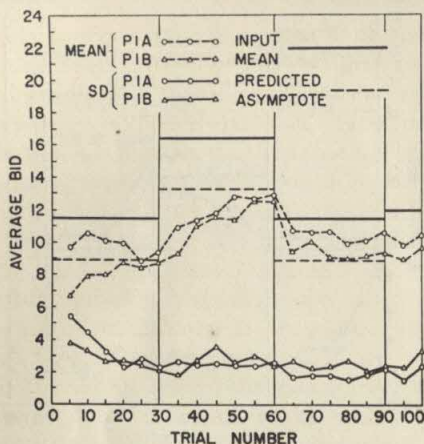ut distributions and the predicted asymptote for each group. The input distribution is the same as that of Fig. 2 for the Guess condition and that of Fig. 4 for the Nonpunish condition.)

satisfied by $b = 9.7$ and for the input 13–20, by $b = 14.1$. These predictions are indicated on Fig. 3 and 4 and are approximately at the observed asymptotes (within one standard error). In contrast, it may be noted that asymptotes of 8 and 13, respectively, would be required if $S$s were either to maximize expected payoff or to minimize expected regret.

*Payoff variable: Punish.*—The data for the Punish groups are shown in Fig. 5 and 6. The results are consistent with those of the other conditions. That is, the mean response levels depend on the mean of the input distribution, and the inter-$S$ SDs are negatively accelerated monotonic decreasing except for slight increases when the input distribution is shifted.

The underbidding regret is here identical to that of the Nonpunish cases. However, in case of an overbid, $S$ loses the amount bid, so that the regret is defined to be the sum of the

input and the response. The asymptotic bid level $b$ should, therefore, satisfy

$$\sum_{j=0}^{[b-1]} (N_j + b)p_j = \sum_{j=[b]}^{\infty} (N_j - b)p_j$$

Values of $b$ that satisfy this equation are 8.9 for the input distribution 8–15, 13.2 for the input 13–20, and 18.1 for the input 18–25. The predicted asymptotes are indicated in Fig. 5 and 6 and appear to be reasonably descriptive of the data. Comparable asymptotic predictions would be 8, 13, and 18, respectively, if $S$s maximized expected payoff or minimized expected regret.

It was noted above that the regret equalization hypothesis evolved from a consideration of the effect of over- or underbidding on a subsequent bid. The conceptualized process produces a decrease in the response level in proportion to the amount of regret experienced as a result of overbidding and an increase in the response level in proportion to the amount of regret due to underbidding. A detailed look at the trial by trial changes in the bid
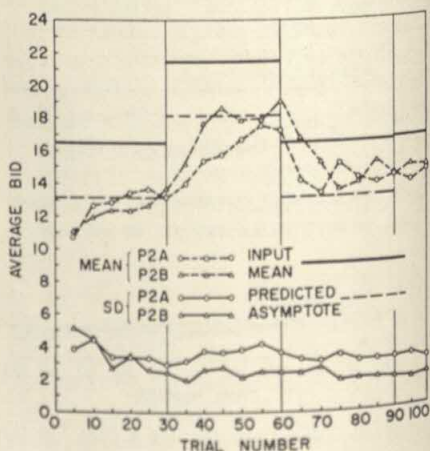


FIG. 6. Means and inter-$S$ SDs for the Punish condition with a *shifting* input distribution having a larger maximum than used elsewhere.

TABLE 2

CORRELATION BETWEEN RESPONSE CHANGE
AND REGRET COMPUTED OVER
30-TRIAL BLOCKS

| Group | Trial Block | | |
|---|---|---|---|
| | 1–30 | 31–60 | 61–90 |
| G1a | .683 | .672 | .593 |
| G1b | .695 | .700 | .577 |
| G2a | .652 | .641 | .643 |
| G2b | .583 | .740 | .738 |
| G3 | .674 | .620 | .644 |
| N1a | .442 | .412 | .266 |
| N1b | .426 | .437 | .359 |
| N2a | .432 | .447 | .462 |
| N2b | .553 | .552 | .505 |
| N3 | .495 | .466 | .411 |
| P1a | .538 | .490 | .491 |
| P1b | .481 | .494 | .568 |
| P2a | .485 | .393 | .424 |
| P2b | .542 | .480 | .518 |

level support this conceptualization. Let regret due to underbidding be arbitrarily defined to be positive and regret due to overbidding be defined to be negative. One can use the product-moment correlation, $r$, between amount of regret and amount of change in the bid level as an index of the extent to which this model of the effect of regret actually describes the data. Such correlations have been computed for each $S$ in each experiment for the three successive blocks of 30 trials. The averages of these $r$'s taken over the group are shown in Table 2.

For all groups, the $r$ is reasonably large on the initial trial block. However, those groups that received the same input distribution throughout the three trial blocks (G1, G3, N1, and N3) seem to show a sharp decrease in the degree of correlation between regret and response change over the three trial blocks. Thus, there is evidence that regret is indeed intimately related to response change,

but that with a constant population distribution for the input, this relation becomes less important. In addition, although no interpretation will be attempted here, it should be noted that all correlations for Guess groups are larger than any for the other groups.

An attempt was also made to provide some additional insight into the values of the correlation coefficients. It was noted that the $SD$s of the bids in a group also decreased to a low value over a series of trials, showing increases only when a change occurred in the input distribution. If the decrease in these $SD$s also implied a decrease in the $SD$ of a series of responses of an individual then that might be sufficient to account for the decrease in correlation of the regret with the change in bid level. Accordingly, such intra-$S$ $SD$s were computed for each $S$ for each block of 30 trials. Averages of these intra-$S$ $SD$s are shown in Table 3. Also given in this table is the $SD$ of

TABLE 3

INTRA-$S$ $SD$s COMPUTED OVER
30-TRIAL BLOCKS

| Group | Trial Block | | |
|---|---|---|---|
| | 1–30 | 31–60 | 61–90 |
| G1a | 4.54 | 2.73 | 2.33 |
| G1b | 3.88 | 2.52 | 2.21 |
| G2a | 3.88 | 3.41 | 2.94 |
| G2b | 3.99 | 3.54 | 3.21 |
| G3 | 4.63 | 2.97 | 2.99 |
| N1a | 3.46 | 1.47 | 1.09 |
| N1b | 2.77 | 1.57 | 1.56 |
| N2a | 2.65 | 2.51 | 2.45 |
| N2b | 3.65 | 2.95 | 2.44 |
| N3 | 3.00 | 1.84 | 1.47 |
| P1a | 3.34 | 2.33 | 1.69 |
| P1b | 2.84 | 2.55 | 2.17 |
| P2a | 2.86 | 2.55 | 1.99 |
| P2b | 2.84 | 2.39 | 2.22 |
| Input | 2.29 | 2.29 | 2.29 |

the input distribution. This value of 2.29 is appreciably larger than the asymptotic $SD$s for at least some of the groups (notably N1a, N1b, and N3—the Nonpunish groups with constant input distributions).

These data show that the Guess groups also tend to exhibit relatively larger intra-$S$ $SD$s, and that appreciable decreases in the intra-$S$ $SD$s occur over the successive trial blocks. This is in line with the decreasing $r$'s of Table 3, since it is clear that the $r$ must perforce be low if the variability in bid level is low. However, the patterns of the decreases of the two sets of data are different. Most of the decrease in correlation occurs only for the groups not getting shifts in the input distribution and it occurs on the last trial block. In contrast, all of the groups show a decrease in intra-$S$ variability, and the decrease takes place primarily on the second trial block.

No explanation is attempted here for these effects. This does not detract from the importance of these effects for learning theory. The data demand that whatever theory or model is used to account for them must also produce these decrements in variability and correlation. In particular, an adequate theory must make the response variability—both intra- and inter-$S$ dependent on the stationarity of the input distribution.

## SUMMARY

Subjects were instructed to ask for some number of "make-believe dollars" (MBDs) or simply to guess a number which $E$ would subsequently present. The payoff to $S$ depended on the relation of $S$'s bid to $E$'s number. Three conditions were used to determine the payoff. In two of these, $S$s were encouraged to bid high, but excessively high bids were punished. In the other condition, over- and underbids were treated symmetrically. A model was constructed which predicts the asymptotic bid level under these conditions to be at a point where the expected regret due to overbidding is equal to the expected regret due to underbidding.

The results indicated: (*a*) The asymptotes of the bids depend on the payoff conditions and the distribution of input numbers as predicted by the model. (*b*) Both the inter- and the intra-$S$ variability decrease over trials except when the distribution of input numbers is changed. (*c*) The increase or decrease in bid level on a trial is highly correlated with the regret associated with the preceding trial.

# SOME EFFECTS OF THE PERCENTAGE OF RELEVANT CUES AND PRESENTATION METHODS ON CONCEPT IDENTIFICATION [1]

MARGARET JEAN PETERSON

*Indiana University*

In order to perform the complex discriminations necessary for successful solution of a concept formation problem, Ss must distinguish between dimensions which are relevant to the solution of the problem and those which are not. Predicting that adaptation of the irrelevant cues and conditioning of the relevant cues would be facilitated by temporally proximate presentation of instances of the relevant dimension, the studies reported herein varied the proximity of relevant instances by presenting all instances relevant to one concept before showing instances representing another concept (homogeneous condition), and by presenting the instances representative of three separate concepts in a mixed sequence (heterogeneous condition).

The percentage of relevant cues per problem was manipulated by using three relevant dimensions and one irrelevant dimension for one problem (75%R); two relevant and two irrelevant for another (50%R); and one relevant and three irrelevant for a third (25%R).

Underwood (1952) emphasized that temporally contiguous presentation of stimuli which are instances related to the same concept should facilitate learning by minimizing the interference effects that might be produced by

interpolated instances of other concepts. Although massed practice has not always been associated with increased efficiency in solving problems (Underwood, 1961), recent demonstration by Cahill and Hovland (1960) of the importance of memory in the acquisition of concepts suggested the prediction that homogeneous presentation would favor faster learning than would heterogeneous presentation of the relevant instances. Further, an interaction between the two variables was predicted; namely, that the advantages of homogeneous presentation would be greater for the 25%R problems than for the 75%R problems.

## EXPERIMENT I

### Method

*Stimulus materials.*—Six three-valued dimensions were used: size (small, medium, large); number of figures (one, two, three); form of the figures (circle, triangle, square); number of lines on the edges of the cards (one, two, three); color (red, blue, green); and position of the figures on the cards (right, middle, left). From this population, dimensions were randomly drawn for the problem subject to the restriction that each dimension be represented equally often as relevant or irrelevant over the entire set of problems. The stimuli were painted with poster paint on white 4 × 6 in. cards. All possible combinations of the irrelevant and the relevant dimensions appeared in a given problem deck. Instances representing the relevant dimensions for the 50% R and 75% R problems were paired using a table of random numbers so that, for example, if color and number of figures were relevant, the single figures were always painted blue; two figures were always red; and three figures, green. Dimensions not used were held constant on all cards within a

## TABLE 1
### Trials to Criterion and Correctly Identified Dimensions as a Function of Conditions and Percentages of Relevant Dimensions: Exp. I and II

| Experiment and Condition | Trials to Criterion | | | | | | | | | Correctly Identified Dimensions | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 25% R | | | 50% R | | | 75% R | | | 25% R | | 50% R | | 75% R |
| | Mean | Mdn. | SD | Mean | Mdn. | SD | Mean | Mdn. | SD | Mean | % | Mean | % | Mean | % |
| Experiment I | | | | | | | | | | | | | | | |
| Heterogeneous | 23.67 | 14.00 | 20.16 | 13.94 | 7.00 | 15.74 | 3.58 | 2.00 | 3.19 | .81 | 81 | .94 | 47 | 1.14 | 38 |
| Homogeneous | 4.69 | 4.00 | 2.92 | 3.56 | 4.00 | 2.03 | 2.08 | 2.00 | 1.38 | .92 | 92 | 1.00 | 50 | 1.31 | 44 |
| Experiment II | | | | | | | | | | | | | | | |
| Proximate (P) | 4.88 | 4.00 | 4.43 | 4.25 | 4.00 | 2.63 | 2.58 | 2.00 | 1.80 | .88 | 88 | 1.00 | 50 | 1.29 | 43 |
| Spaced, Filled ($S_D$) | 7.83 | 7.00 | 4.12 | 4.44 | 6.00 | 3.37 | 3.33 | 3.00 | 2.05 | .92 | 92 | .92 | 46 | 1.17 | 39 |
| Spaced, Unfilled ($S_{\bar{D}}$) | 3.83 | 3.00 | 2.67 | 3.00 | 2.00 | 1.47 | 1.75 | 1.50 | 1.30 | .92 | 92 | 1.00 | 50 | 1.25 | 42 |

problem deck, e.g., when color was not used as either a relevant or an irrelevant dimension, all stimuli were painted gray. The 3 × 2 factorial design consisted of the three percentages of relevant dimensions and the two methods of presentation. Each problem was presented under both methods of presentation. Each S solved all six problems whose sequential order of appearance had been determined by a Latin square design such that each problem appeared equally often in every ordinal position.

*Apparatus.*—Cards were viewed through a one-way vision mirror mounted in a large black screen. The S indicated his response by pressing one of three telegraph keys. A reinforcing light was placed immediately above each key.

*Subjects and procedure.*—The Ss were 36 students from introductory psychology courses at Indiana University. Experimental participation was a course requirement. After instructions defining S's task, including an enumeration of the possible bases for discrimination, a set of practice cards which had the letters A, B, or C on them was shown to S until he had responded correctly three times. Homogeneous presentation consisted of showing cards representative of one concept until S had correctly identified the cards three consecutive times. Then representations of the second concept were shown to the same criterion, followed by the presentation of the third concept. *Concept* refers to the relevant stimulus characteristics of cards associated with *one* of the response keys so that Ss were said to learn three concepts to define a problem. If color were the relevant dimension, learning Response A to red represented one concept; Response B to blue was the

second; and Response C to green, the third. In heterogeneous presentation instances of the three concepts were assigned randomly with the restriction that one instance of the three concepts appear in each block of three cards. The Ss were run to a criterion of 9 consecutive correct responses or until 54 trials had been completed. Each S was then queried about the basis for solution of the problem before going on to the next.

## Results and Discussion

Generally fewer responses were required to reach the criterion following homogeneous presentation than following heterogeneous presentation. Higher percentage relevant problems were learned more rapidly than low percentage ones (Table 1).

A Friedman two-way analysis of variance (Siegel, 1956, p. 166), employed because of the heterogeneity of variance, yielded a significant $\chi^2_r$ of 18.35 ($df = 2$, $P < .001$). The chi square for the three levels of percentage of relevant dimensions with heterogeneous presentation was also significant ($\chi^2_r = 27.56$, $df = 2$, $P < .001$). The $z$ conversions for the sign test (Siegel, 1956, p. 72) used to assess the differences between presentation conditions within the 25%R, 50%R, and 75%R conditions were 3.04 for 25%R ($P = .0012$), 2.70 for 50%R

$(P = .0035)$, and 1.33 for 75%R $(P = .0934)$. The predicted interaction was found: homogeneous presentation had a greater facilitating effect for the low percentage of relevant dimensions than for higher ones.

Hull (1920) in his classical study of concept identification using Chinese characters noted that Ss were not necessarily able to define verbally the property common to a specific concept even though they could assign stimuli to concepts correctly. In contrast, Bourne and Haygood (1959) reported that their Ss were almost always able to label the correct dimensions, even when more than one dimension was relevant for a particular problem. Table 1 contains an analysis of verbal identification in the present experiment. The mean number of correctly identified dimensions increased significantly from the heterogeneous to the homogeneous presentation ($z$ transformation of sign test = 2.94, $P = .0016$) and increased as the percentage of relevant dimensions increased (Friedman $\chi^2_r = 14.59$, $df = 2$, $P < .001$). The proportions of the number of correct identifications relative to the total number of correct dimensions that could have been named demonstrated clearly that the majority of Ss were reporting only one dimension, even when additional dimensions could have been used: 2 of the 36 Ss identified both correct dimensions for the 50%R problems; 16 of the 36 Ss identified two of the three correct dimensions for the 75%R problems; but no S identified all three. The numbers of correct assignments of concept instances to the response keys were not reported, since the $\rho$'s with the number of correct identifications of the dimensions ranged from .82 for the 25%R

problems to .96 for the 75%R problems.

## EXPERIMENT II

The superiority of homogeneous presentation in Exp. I may have resulted from the closer proximity of instances of a given concept in that condition. Another possibility is that the absence of interference from presentation of instances of other concepts permitted faster learning. In Exp. II the problems were presented using a homogeneous sequence while preserving the exact temporal ordering of the instances in the related heterogeneous condition of Exp. I. The intervals were filled with a digit cancellation task for one group and left unfilled for another. The control Ss learned the problems using the homogeneous condition of Exp. I.

### Method

Both of the two experimental conditions, temporally proximate (P) and spaced (S), used the homogeneous sequence of presentation of Exp. I. Variations in temporal separation of the concept instances distinguished the two conditions. Condition P was identical with the homogeneous condition of Exp. I. In Cond. S, instances of the first concept were shown using the temporal intervals which existed in the heterogeneous problem of the same percentage of relevant dimensions, but the intervals were either filled with digit cancellation ($S_D$) or left unfilled ($S_{\bar{D}}$). During the unfilled intervals S sat silently in front of the darkened aperture. Then instances of the second concept arranged to simulate its heterogeneous problem presentation sequence were shown followed by the simulation of the third. For all groups, instances of one concept were presented until S had emitted three consecutive correct responses before instances of the second were presented.

The Ss, 36 students from introductory psychology courses at Indiana University who had not participated in similar experiments, were assigned randomly with the restriction that an equal number of Ss experience each condition. The remainder of the experimental procedure was identical to that of Exp. I.

## Results and Discussion

Application of a Kruskal-Wallis one-way analysis of variance of trials to criterion (Siegel, 1956, p. 184) yielded an $H$ of 6.93 comparing the three conditions of presentation (Table 1) for the 25%R problems which, with 2 $df$, was significant between the .02 and .05 levels; the comparable $H$s for the 50%R problems (4.66) and the 75%R problems (2.66) were associated with probabilities greater than .05, both with 2 $df$. Differences between the temporally proximate presentation of the 25%R problem and either of the spaced methods of presentation were not statistically reliable; however, the use of digit cancellation did significantly increase the number of trials required for solution relative to the spaced condition with an unfilled interval ($P = .018$ using the median test, Siegel, 1956, p. 111). Apparently, lengthening the interval between instances of a concept did not in itself significantly slow learning. Rather, interference from instances displaying another concept or the introduction of another task such as digit cancellation appeared to retard learning, particularly with problems characterized by a low percentage of relevant cues.

Results of $Ss$' identification of the correct dimensions reflected the trends shown in Exp. I. The higher the percentage of relevant cues the more frequently $Ss$ were able to label at least one correct dimension, although the different manipulations of the homogeneous condition in Exp. II were not portrayed in these data. Again, few $Ss$ identified more than one correct dimension. No $Ss$ reported two for the 50%R problems, nine reports of two correct dimensions were given for the 75%R problems, and two

reports of three dimensions were made. Differences in the mean number of trials to criterion and in the mean number of correctly identified dimensions between comparable conditions of Exp. I and II were not statistically significant, both chi squares being less than 1.

## DISCUSSION

Because lengthening the interval between presentations of instances of the same concept did not have a significant effect upon the learning of concepts, the efficacy of homogeneous presentation did not appear to reflect massing of practice, per se. Introduction of conditions which would be expected to increase the likelihood of some kind of interference such as the heterogeneous method of presentation or the digit cancellation task was associated with slower learning of the concepts, particularly when the concepts to be learned contained a low percentage of relevant dimensions. It is possible that when a high percentage of the dimensions were relevant the problems were learned so rapidly that these factors became relatively unimportant or exerted an influence too transitory to be reflected in the response measures employed. Furthermore, an unpublished replication of Exp. I, using different dimensions to constitute the problems, yielded almost identical results.

Another observation was the infrequent identification of more than one correct dimension even when additional dimensions were available and each $S$ had been told what dimensions might be used. The assumption might be made that as relevant dimensions were added, the stimulus pool from which $S$ sampled would have increased so that $Ss$ were actively selecting from a larger population of relevant cues for the problems of higher percentage of relevant cues then for problems of lower percentage. More in accord with the data would be the interpretation that over a group of $Ss$ the probabilities increased that each $S$ identify at least one correct dimension

without necessarily having been able to report the presence of other relevant dimensions.

## SUMMARY

Two experiments examined the effects of the variation of the percentage of relevant dimensions and the method of presentation of concept instances on rate of concept identification. Problems consisting of 25%, 50%, and 75% relevant cues were combined factorially with four different dimensions. Instances of one concept were presented until the criterion of learning had been achieved, then instances of the second concept were presented followed by the third for the homogeneous condition. In the heterogeneous condition, instances of the three concepts were presented in a random sequence. The predictions that the number of responses prior to criterion would be inversely related both to the percentage of relevant cues and to the temporal proximity of the instances associated with a given response were supported. Homogeneous presentation was more advantageous with 25% R than with 50% R and 75% R. Experiment II demonstrated that the lesser efficiency of heterogeneous presentation was not a function of the greater temporal intervals occurring between instances of the same concept, but rather of interference effects from other concepts, at least with 25% R problems.

Analyses of correctly identified dimensions suggested an interaction effect between the percentage of relevant cues and the method of presentation. Few Ss reported the presence of more than one relevant dimension for the problems with two or three completely redundant relevant dimensions.

## REFERENCES

BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, 58, 232–238.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *J. exp. Psychol.*, 1960, 59, 137–144.

HULL, C. L. Quantitative aspects of the evaluation of concepts. *Psychol. Monogr.*, 1920, 28 (1, Whole No. 123).

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

UNDERWOOD, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, 59, 209–220.

UNDERWOOD, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, 68, 229–247.

# EASE OF CONCEPT ATTAINMENT AS A FUNCTION OF ASSOCIATIVE RANK [1]

SARNOFF A. MEDNICK[2]     AND     SHARON HALPERN

*University of Michigan*        *University of California, Berkeley*

Underwood (1952) has suggested a method for the study of concept formation which assumes that the attainment of a concept calls for the perception of a relationship between concept instances. The perception of this relationship, in part, depends on the probability of the occurrence of the relevant associative response to the concept instances. This probability is termed response dominance. The mean response dominance of all instances representing the concept is termed dominance level. Underwood and Richardson (1956b) have shown that the ease of attainment of a concept is directly related to its dominance level.

This study explores a methodological variable which determines ease of concept attainment. The variable under investigation is the rank position of the concept response in the associative hierarchy of the concept instance. To the concept instance BELLY the sensory associate ROUND is of Rank 1 with a dominance level of 43% (Underwood & Richardson, 1956a). The sensory associate SOFT is of Rank 2 with a dominance level of 24%. To the concept instance,

PAIL the sensory associate METALLIC is of Rank 1 and has dominance level of 24%. While METALLIC to PAIL and SOFT to BELLY are equal in response dominance they vary in their positions in their respective associative hierarchies: METALLIC is of Rank 1; while SOFT is of Rank 2.

This experiment compares ease of attainment of concepts as a function of the rank position of the concepts in the associative hierarchy of the concept instances. For reasons developed below it is predicted that first ranking concepts will be attained in fewer trials and with fewer errors than will second ranking concepts.

## METHOD

*Lists.*—The words used were concrete nouns selected from a list of 213 nouns for which Underwood and Richardson (1956a) have ascertained the dominance level of various responses. As can be seen in Table 1, four groups of instances were assembled: associative Rank 1 (AR 1) WHITE, AR 1 ROUND, associative Rank 2 (AR2) WHITE, and AR 2 ROUND. This was necessary because of the possibility that the concepts might differ in difficulty or that concept difficulty might interact with associative rank. List 1 consisted of AR 1 ROUND and AR2 WHITE while List 2 contained the other two concepts. A buffer concept "LONG" included in both lists (EEL, BEAK, ALLEY, CUCUMBER) was used to make it more difficult for Ss to attain the concepts by elimination procedures. As is shown in Table 1, the mean dominance levels of concepts were kept nearly constant. In constructing the AR 2 concepts care was taken to avoid having the concept instances elicit a common first ranking response.

The instances were presented to S in three random orders. The same three orders were used for both lists with positions occupied by

TABLE 1

CONCEPT LISTS WITH ASSOCIATIVE RANK, RESPONSE DOMINANCE, AND MEAN
DOMINANCE LEVEL INDICATED

| Associative Rank in Hierarchy | List I | | | List II | | |
|---|---|---|---|---|---|---|
| | Noun | Response Dominance | Concept and Mean Dominance | Noun | Response Dominance | Concept and Mean Dominance |
| 1 | POT<br>EYE<br>DIME<br>GRAPE | 29%<br>32%<br>30%<br>18% | Round<br>27% | HOSPITAL<br>ENAMEL<br>GOAT<br>BREAD | 32%<br>28%<br>29%<br>35% | White<br>31% |
| 2 | FROST<br>GARDENIA<br>LARD<br>BONE | 34%<br>28%<br>27%<br>34% | White<br>31% | BADGE<br>PILL<br>WAIST<br>CAPSULE | 21%<br>28%<br>24%<br>22% | Round<br>24% |

AR 1 instances in List 1 being occupied by AR 2 instances in List 2. The positions of the buffer terms were not changed between lists, although they were randomized in the three orders that were used.

*Subjects.*—The Ss were 30 undergraduate paid volunteers. The 15 men and 15 women were divided as equally as possible between the two lists.

*Procedure.*—The lists were presented at a 4-sec. rate on a Gerbrand's type memory drum. A 12-sec. interval occurred between presentations of the lists.

The Ss were informed that the list contained 12 words that could be placed in three groups of 4 words each, and that all of the 4 words in each of these groups could be described by the same adjective. The Ss were required to respond to each word. The task was continued to a criterion of one perfect trial, or terminated at 20 trials. A more complete discussion of the materials and procedure may be found elsewhere (Freedman & Mednick, 1958).

## RESULTS AND DISCUSSION

The data which were subjected to analysis were the number of trials to one perfect trial on a concept (this meant giving the correct concept response to all four instances of a concept on the same trial) and the number of errors made on each concept in the entire course of the experiment. As noted above, the buffer concept LONG was omitted from this analysis. AR 1 concepts (List 1, WHITE, List 2, ROUND) were compared with AR2 concepts (List 1, ROUND, List 2, WHITE). Two List 2 Ss failed to solve any concept and were dropped from further analysis.

The AR 1 concepts were attained earlier. The mean number of trials taken to solve each concept was 5.90 for the AR 1 concepts and 8.09 for the AR 2 concepts, a significant difference ($t = 2.96$, $df = 27$, $P < .01$). The mean number of errors on the AR 1 and AR 2 concepts were 12.17 and 18.14, respectively. This difference was significant ($t = 2.09$, $df = 27$, $P < .05$).

The results are intuitively satisfying but detailed analysis of their interpretation is somewhat intricate. We have found that AR 1 concepts are attained more easily than AR 2 concepts despite the fact that their dominance levels are equal. Above, we have referred to these ranks as indicating a position in an associative hierarchy. However, this hierarchy is, in a sense, a figment. The norms which provide us with dominance levels and ranks are based on Ss giving single sensory associates to each of 213 nouns. Actually, while WHITE is a second ranking response to GARDENIA and a first ranking response to ENAMEL

28% of the norm group gave WHITE as their *first and only response* to these nouns, and in both cases, 72% gave some other response. Thus, when we refer to these concept responses as occupying positions in an *S*'s associative hierarchy we are making the implicit assumption that the associative hierarchy produced by collating the group's single responses is reflected to a large extent, in each individual. In other words, we are assuming that everyone has just about the same basic associative hierarchy; the fact that we get variation in single response norms we would then attribute to momentary fluctuations in associative strength. Thus, if Underwood and Richardson (1956a), had asked their *S*s to give more than one response to each noun, a large proportion of the 72% that did not give WHITE as their first response to ENAMEL would have given it as their second or third response. If this situation were applied to the present experiment then the superiority of the AR 1 concepts would be understandable. The AR 2 concept responses occupy an inferior position (relative to the AR 1 concept responses) in almost everyone's associative hierarchy. This means the AR 1 concept responses would be elicited earlier in the course of the experiment. This experiment may then be seen as supporting this assumption of homogeneity of hierarchies. Research on word associations (Cofer, 1958; Rosen & Russell, 1957) has contributed consider-able support to this same assumption in another context.

## Summary

Thirty *S*s were presented with lists of 12 nouns and instructed to discover into what three groups the nouns could be divided and what adjective could describe each group. The lists consisted of concepts of equal levels of dominance; the position of the concept responses in the associative hierarchy was manipulated. The concepts having higher rank position in the associative hierarchy were attained more quickly and with fewer errors.

## References

COFER, C. N. Comparison of word associations obtained by the methods of discrete single word and continued association. *Psychol. Rep.*, 1958, **4**, 507–510.

FREEDMAN, J. L., & MEDNICK, S. A. Ease of attainment of concepts as a function of response dominance variance. *J. exp. Psychol.*, 1958, **55**, 463–466.

ROSEN, E., & RUSSELL, W. A. Frequency characteristics of successive word association. *Amer. J. Psychol.*, 1957, **70**, 120–122.

UNDERWOOD, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–220.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95. (a)

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238. (b)

# CONCEPT IDENTIFICATION UNDER MISINFORMATIVE AND SUBSEQUENT INFORMATIVE FEEDBACK CONDITIONS [1]

## WALTER J. JOHANNSEN

*Veterans Administration Center, Wood, Wisconsin*

Recent research on human concept identification has aimed at delineating the effects of feedback class on attainment rate. In particular a recent study by Pishkin (1960), using misinformative feedback (MF), reveals striking decrement when even small percentages of erroneous, task relevant feedback information are inserted into schedules of informative or correct feedback (IF). The present study seeks to extend Pishkin's results by assessing the effect of MF on subsequent concept identification under conditions of 100% IF.

Appropriate design makes possible the concurrent examination of another, associated question. Pishkin found probability matching behavior in his concept identification study, as did Goodnow and Postman (1955) in a study using MF by implication. On the other hand, Morin (1955), who made use of MF in a simpler learning situation, was unable to demonstrate an adequate match in his data. He suggested that the failure of the obtained curves to approach an asymptote was a factor in his results. The present experiment is designed to circumvent this problem by carrying performance under MF/IF to a point

where asymptote is more nearly approximated.

A final interest of this paper is the description of acquisition curves under MF/IF. Pishkin's analysis, related to the Restle (1955) discrimination learning model, is unconcerned with the nature of the attainment process. Yet Morin's trend analysis of his data suggests a more complex process than that typical of the probability matching studies. It is of interest to determine whether these findings can be replicated in data derived from a more difficult task.

## METHOD

*Experimental conditions.*—The Ss were randomly assigned to one of four MF/IF conditions and one of three task complexity conditions. The percentages of MF/IF employed were 0:100, 12.5:87.5, 25:75, and 37.5:62.5. Task complexity was simultaneously manipulated by varying the number of dimensions irrelevant to problem solution while holding constant the number of relevant dimensions. A single dimension was relevant for all conditions, and either 1, 3, or 6 dimensions were irrelevant. The design therefore describes a 3 × 4 orthogonal plot. Ten Ss were tested in each of the 12 resulting cells.

*Subjects.*—The Ss were 124 sophomore psychology students attending the University of Wisconsin. All had volunteered for the experiment in order to gain credit applicable to their class grades. Four Ss were eliminated for failure to comply with instructions.

*Apparatus and procedure.*—Stimuli consisted of geometric figures drawn on 3 × 5 in. cards. Each card was inscribed with a single figure. Figures varied according to the following dimensions and values within dimensions: form (rectangle-triangle), size (large-small), location (center-right of center),

## TABLE 1

Duncan Range Analysis: Mean
Performance Scores during
MF/IF Trials

| Condition (MF%:DI) | Mean | SD |
|---|---|---|
| 0:1 | 198.8 | .83 |
| 0:3 | 194.4 | 5.02 |
| 0:6 | 186.6 | 9.29 |
| 12.5:1 | 173.8 | 22.26 |
| 25:1 | 142.5 | 33.89 |
| 12.5:3 | 142.0 | 28.99 |
| 12.5:6 | 141.4 | 32.79 |
| 25:3 | 119.6 | 15.20 |
| 37.5:1 | 105.3 | 10.39 |
| 37.5:3 | 102.7 | 10.60 |
| 25:6 | 100.9 | 5.26 |
| 37.5:6 | 96.3 | 13.41 |

Note.—Means joined by vertical line do not differ
significantly; means not so joined are significantly
different (P < .05).

position (vertical-horizontal), figure color
(black-blue), ground color (red-white), dot
within figure (presence-absence).

The apparatus consisted of a 20 × 36 in.
flat-black panel mounted vertically on a table
of normal height, which served to separate S
from E. A 3 × 5 in. aperture was cut into
the center of the panel slightly below eye
level. Stimuli were manually inserted into
this aperture from the rear.

Two 7.5-w. bulbs, one red and one white,
were mounted side by side in sockets set 6 in.
apart, 8 in. above and to either side of the
presentation aperture. These lights served
as feedback signals and were controlled by
E, using two Western Union telegraph keys.

Instructions were read to S, informing him
that he was to take part in a concept identi-
fication experiment, and that his task would
involve the classification of cards placed
before him. Specifically, S was told to label
each card either A or B, with each A card
having something in common and each B
card having something in common. The
flashing of a white light would indicate a
correct response and a red light an incorrect
response. No references were made to the
presence of MF. Groups serving under the
different dimensions-irrelevant (DI) condi-
tions were read supplementary instructions
in accordance with Hovland's (1952) pro-
cedure, in which Ss are informed of the range
of values and dimensions available to them.
For all groups Category A constituted a
vertical figure and B a horizontal figure.
Following these instructions questions were

answered with a paraphrase of the original
instructions.

A schedule of MF/IF was developed for
each condition. The occurrence of MF was
randomized within each block of 10 trials,
but with each block receiving approximately
the same number of MF trials. All Ss within
a feedback group performed under the same
schedule.

All experimental Ss received 200 trials
under MF/IF conditions and were then
shifted to a 100% IF schedule until a criterion
of 10 successive correct responses was
achieved. Control Ss who usually made long
runs of correct responses in less than 100 trials
were terminated after 20 successive correct
responses and, for purposes of analysis, were
credited with an additional number of correct
responses equal to the difference between 200
and the number of the terminal trial.

On a given trial E randomly selected a
stimulus card from the shuffled pack before
him and placed it in the presentation aperture
in front of S. After S responded verbally, E
recorded the classification of the card (A or B)
and whether S had responded correctly or
incorrectly. After reference to the MF/IF
schedule, E determined whether MF or IF
was to be administered on that trial and
pressed one of the two keys to signal feedback.
Average time per trial was approximately
10 sec.

## RESULTS AND DISCUSSION

*Performance under MF/IF condi-
tions.*—Prior to examination of the
acquisition process, a preliminary
analysis of variance was performed on
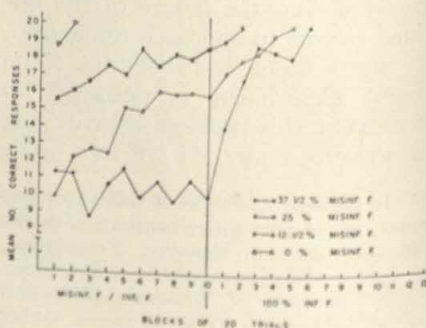the mean number of correct responses



Fig. 1. Mean number correct responses
per 20-trial block under MF/IF and sub-
sequent IF conditions. (Parameter is MF
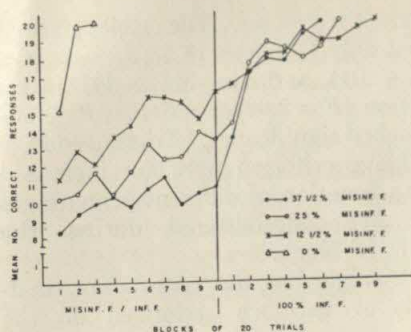percentage. One irrelevant dimension.)

FIG. 2. Mean number correct responses per 20-trial block under MF/IF and subsequent IF conditions. (Parameter is MF percentage. Three irrelevant dimensions.)

during the MF/IF block, but also including control conditions. The results demonstrate significant differences as a function of DI ($F = 14.60$, $df = 2/108$, $P < .01$), of MF% ($F = 668.20$, $df = 3/108$, $P < .01$) and of the interaction between the two variables ($F = 6.98$, $df = 6/108$, $P < .01$). A supplementary Duncan range analysis, reported in Table 1, indicates the ordering of means and the position of significant differences dividing the cells.

As anticipated, the increasing number of DI is related to poorer performance, in agreement with earlier research (Archer, Bourne, & Brown, 1955). The extremely large $F$ ratio ascribed to MF% is partially a function of bias introduced by inclusion of control conditions where optimal performance was reached early in training. Examination of the Duncan range results indicates that the number of correct responses diminishes regularly with increasing stimulus difficulty and MF%, although a few inversions of order exist.

In order to analyze the acquisition process under MF/IF conditions, a trend test (Grant, 1956) was performed on the group scores, with the number of correct responses per 20-

trial block providing the raw data. Performance curves obtained from each of the 12 conditions appear in Fig. 1–3. Because the small cell frequencies and the limitations on the range of possible cell scores yielded truncated distributions, an arc-sine transformation was performed (Snedecor, 1946, p. 445) and analysis conducted on the transformed data. Significant differences between group means occur as a function of MF% ($F = 38.53$, $df = 2/81$, $P < .01$) and DI ($F = 10.33$, $df = 2/81$, $P < .01$) when control data are omitted. Elimination of the control cells reduces the MF $\times$ DI interaction to the extent that it is no longer significant. The overall acquisition curve appears complex, consisting of significant linear ($F = 73.13$, $df = 1/81$, $P < .01$), quadratic ($F = 31.56$, $df = 1/81$, $P < .01$), and cubic ($F = 10.33$, $df = 1/81$, $P < .01$) components. Group differences occur only in the slope of the different MF curves ($F = 91.25$, $df = 2/81$, $P < .01$). These results are consistent with Morin's observation of curve components of a higher order than quadratic.

*Performance under subsequent 100% IF conditions.*—Analysis of subsequent learning under 100% IF condi-
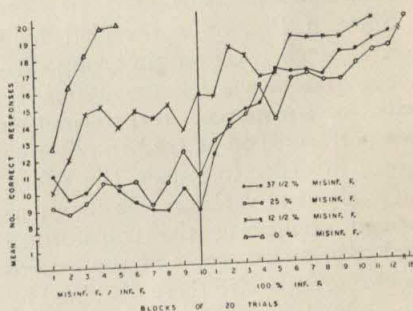


FIG. 3. Mean number correct responses per 20-trial block under MF/IF and subsequent IF conditions. (Parameter is MF percentage. Six irrelevant dimensions.)

tions involved several procedural problems. A few Ss serving in the 12.5% MF cells achieved levels of errorless performance during the last block of MF/IF trials. To drop these Ss from the succeeding IF series would have resulted in a sampling bias, in the sense that the "better learners" would be eliminated from the simpler conditions and retained in the more difficult conditions. The alternative, which was adopted, was to retain these Ss and require achievement of the same criterion as the other Ss. The net effect would be slightly to enhance the probability of a spuriously significant difference between MF groups on learning under 100% IF conditions.

Control of differential group attainment under MF/IF was also needed. It was reasoned that performance differences under 100% IF was a joint function of DI, prior learning, and the residual effects of the terminated MF. Thus, analysis of variance dealing with trials to criterion under 100% IF would yield an apportioning of variance attributable to the effect of experimental variables combined with the effect of previous learning. An analysis of covariance, partialing out the effect of earlier learning, would allow evaluation of the experimental variables alone.

Since a Pearson $r$ revealed high correlation between within-group variances and means for the 100% IF cells, a square-root transformation was performed on all trial-to-criterion scores in order to reduce the effect. An analysis of variance was then performed comparing the transformed trial-to-criterion scores of the experimental groups with those obtained by the control groups in order to determine whether the 200 MF/IF trials had acted to increase the number of trials to achieve a level of 10 successive

correct responses. The results showed that only DI ($F = 17.58$, $df = 2/108$, $P < .01$) and the MF $\times$ DI interaction ($F = 2.92$, $df = 6/108$, $P < .05$) reached significance. Trials needed to achieve criterion were not increased as a function of different percentages of MF administered during the MF/IF block.

Disregarding the control data, analysis of variance performed on the transformed trials-to-criterion scores of the experimental cells provides a similar picture. Here only the differences between DI groups achieve significance ($F = 10.31$, $df = 2/80$, $P < .01$). However, if the effect of prior learning is partialed out (in terms of terminal level of performance during the last two blocks of MF/IF), the MF variable becomes significant ($F = 6.78$, $df = 2/80$, $P < .01$) and the effect of DI is sharply diminished. Thus performance under subsequent IF conditions is affected by the terminated MF trials, but the effect is complex and requires further explication.

*Probability matching.*—Probability matching behavior was evaluated by subtracting each S's attained number of correct responses on the terminal 40 MF/IF scores from the theoretically expected scores. These difference scores were evaluated by separate $t$ tests for each cell. The results present a complex picture. Adequate matching was noted on five of the nine experimental cells: the three 12.5% MF groups, the 25% MF-1 DI, and the 25% MF-3 DI conditions. Significant negative deviations from probability matching were noted on all three 37.5% MF cells and a significant positive deviation on the 25% MF-1 DI cell. The breakdown of probability matching in the 37.5% MF conditions is noteworthy but not unique, since a

similar phenomenon occurs in Pishkin's most difficult MF conditions. One possible explanation resides in the fact that Ss' use of response patterns which yield positive reinforcement on 67.5% of the trials are not distinguishably more effective than use of patterns which are successful on 50%, a level which could be reached by randomly responding.

## Summary

The effect of percentage of misinformative feedback (MF: 0, 12.5, 25, 37.5%) and the number of dimensions irrelevant to solution (DI: 1, 3, 6) on acquisition in concept identification and on subsequent performance under 100% informative feedback (IF) were investigated. A total of 120 Ss served, with 90 experimental Ss being administered 200 MF/IF trials, then shifted to 100% IF until criterion was reached.

The results were: (a) Under MF/IF conditions significant differences occurred as a function of MF, DI, and MF × DI, with increasing MF and DI leading to poorer performance. (b) Trend analysis on blocks of trials under MF/IF revealed a curve composed of significant liner, quadratic, and cubic components; the linear component was significantly affected by MF%. (c) Subsequent 100% IF learning was significantly affected by DI; inclusion of control Ss in the analysis lead to MF × DI achieving significance. (d) Analysis of covariance on the 100% IF data, partialing out the effect of prior learning, revealed only a significant MF effect. (e) Probability matching appeared in five of nine MF/IF cells.

## References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, 49, 153–164.

Goodnow, J. J., & Postman, L. Probability learning in a problem solving situation. *J. exp. Psychol.*, 1955, 49, 16–22.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141–154.

Hovland, C. I. A communication analysis of concept learning. *Psychol. Rev.*, 1952, 49, 461–472.

Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, 49, 343–351.

Pishkin, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *J. exp. Psychol.*, 1960, 59, 371–378.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, 62, 11–19.

Snedecor, G. W. *Statistical methods.* (4th ed.) Ames, Iowa: State Coll. Press, 1946.

# RESISTANCE TO EXTINCTION AS A JOINT FUNCTION OF REWARD MAGNITUDE AND THE SPACING OF EXTINCTION TRIALS [1]

## WINFRED F. HILL AND NORMAN E. SPEAR

### *Northwestern University*

The effect of reward magnitude on resistance to extinction is an unsettled question, even for that subset of studies in which the independent variable is the weight of food on a continuous reinforcement schedule and the dependent variable is the running speed of rats. Metzger, Cotton, and Lewis (1957) and Zeaman (1949) found that a larger reward gave faster running early in extinction, with the group curves tending to converge as extinction proceeded. This is what would be expected if in extinction $K$ (Hull, 1951; Spence, 1956) adjusts to the absence of reward from different levels. On the other hand, Armus (1959) and Hulse (1958) found faster running throughout extinction after a smaller reward. This might reflect a contrast or depression effect for the large reward group

The most prominent difference in procedure between these two sets of studies was in the distribution of the extinction trials. Metzger, Cotton, and Lewis and Zeaman gave massed extinction, whereas Armus and Hulse gave spaced extinction. The present experiment is a test of the hypothesis that reward magnitude and spacing of extinction trials will interact within a single experiment. If confirmed, this relationship would be of considerable significance for the interpretation of extinction.

## METHOD

The $S$s were 64 experimentally naive female albino rats of the Sprague-Dawley strain, 74 days old at the beginning of training. Training and extinction took place in an enclosed runway previously described by Lewis (1956).

Each $S$ received six daily 3-min. sessions of prehandling, the last session 48 hr. prior to the beginning of experimental training. During each session $S$ was allowed to explore a large unpainted wooden box, presented with four of the pellets later to serve as reward, and picked up and replaced at least five times by $E$. A once-daily feeding schedule began on the first day of prehandling and was maintained throughout experimental training. The ration was 10 gm. of finely ground Purina lab chow and was presented 50 to 60 min. after the start of prehandling or experimental training.

All $S$s received 25 trials of acquisition, 5 per day, and 20 trials of extinction beginning on the sixth day. During both acquisition and extinction, $S$ was confined in the goal box for a minimum of 15 sec. or until all pellets were consumed (maximum of 4 min.). Between trials on the same day, $S$ was confined in its home cage, with water available, for 20 sec. During extinction, the food cup was removed from the goal box.

Differential training was introduced by way of a 2 × 2 factorial design, varying the number of .045-gm. Noyes pellets given as reward during acquisition (four pellets or one) and the intertrial interval during extinction (20 sec. or 24 hr.). Thus the four experimental groups of 16 $S$s each may be designated according to extinction spacing or massing and according to acquisition magnitude as Sp-4, Sp-1, M-4, and M-1.

## RESULTS AND DISCUSSION

*Acquisition.*—Curves of acquisition speed are shown in Fig. 1. Double classification analysis of variance on the mean speeds for the last five trials
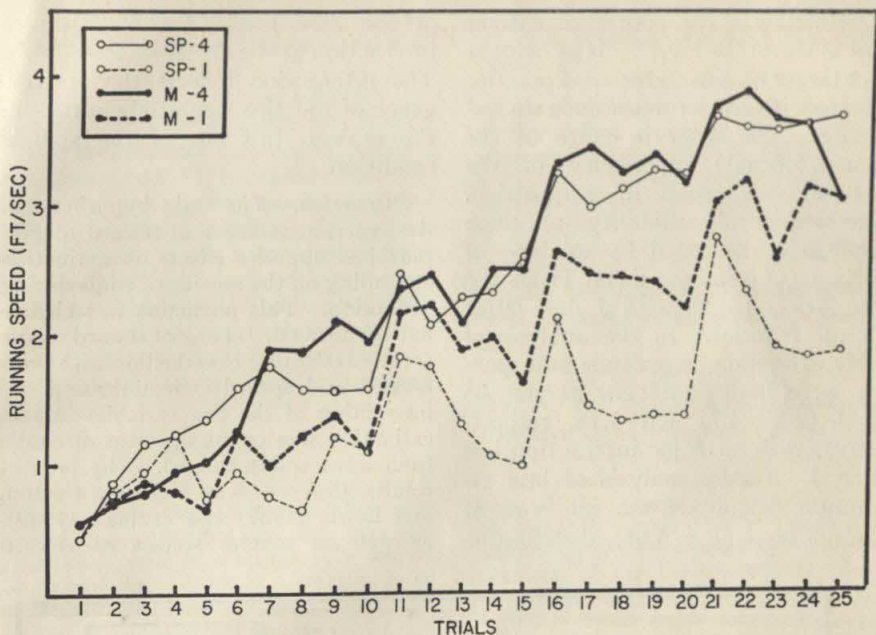
FIG. 1. Mean speeds in acquisition (five trials a day) for two magnitudes of reward and for Ss subsequently receiving massed or spaced extinction.

confirms the superiority of the four-pellet condition ($F = 18.90$, $df = 1/60$, $P < .001$). The dummy distribution variable and the interaction are both nonsignificant ($Fs = 2.91$ and $3.44$, respectively), indicating that the two spacing groups were roughly equivalent before thes pacing variable was introduced.

The curves show a tendency for the greatest increases in speed to come between the end of 1 day and the beginning of the next. This reminiscence was more marked in the later stages of learning and in the one-pellet groups, combined under these conditions with a marked within-days decrement in speed. To quantify this reminiscence effect, a score was computed for each S on each day by subtracting the main gain in speed between each trial and the next from the gain in speed between the last trial of the previous day and the first trial of the day in question. When this score is averaged over the 4 days of acquisition (excluding the first day, for which it cannot be computed), the mean is significantly positive at the .001 level for both the four-pellet (Sp-4 plus M-4) and the one-pellet (Sp-1 plus M-1) groups ($t$'s = 4.85 and 6.75, respectively, for the difference from zero). This indicates that the trial-to-trial gain was greater over the 1-day interval than over the 20-sec. interval. A trend analysis showed the overall mean to be significantly higher at the .05 level for the one-pellet than for the four-pellet group ($F = 3.98$, $df = 1/62$). The increase over trials yielded a significant $F$ of 3.98 ($df = 3/186$, $P = .01$) but one which is not quite significant for the 1 and 62 $df$ recommended as conservative by Geisser and Greenhouse (1958). The $F$ for Group × Trend interaction was less than 1.

*Extinction.*—The course of extinction is shown in Fig. 2. It is evident that larger reward and spaced practice resulted in greater resistance to extinction, the latter in spite of the (nonsignificant) superiority of the to-be-massed group in acquisition. The statistical reliability of these findings is confirmed by analysis of variance of mean speeds on Trials 2–6 and Trials 16–20, with 1 and 60 *df* for all *F* ratios. In the analysis of early extinction, magnitude and spacing were both significant at the .05 level (*F*s = 4.71 and 5.18, respectively), with an *F* for interaction less than 1. In the analysis of late extinction, magnitude was significant at the .05 level (*F* = 6.68), distribution

at the .001 level (*F* = 26.28), and interaction at the .01 level (*F* = 7.45). The interaction reflects the convergence of the two magnitude curves in the massed but not in the spaced condition.

*Discussion.*—The main hypothesis of the experiment was that reward magnitude has opposite effects on extinction depending on the spacing of trials during extinction. This prediction was clearly not confirmed. Larger reward gave greater resistance to extinction with both massed and spaced extinction, and the interaction of the two variables late in extinction was in the opposite direction from what was predicted. The present results thus confirm Metzger, Cotton, and Lewis (1957) and Zeaman (1949), as well as several studies of reward
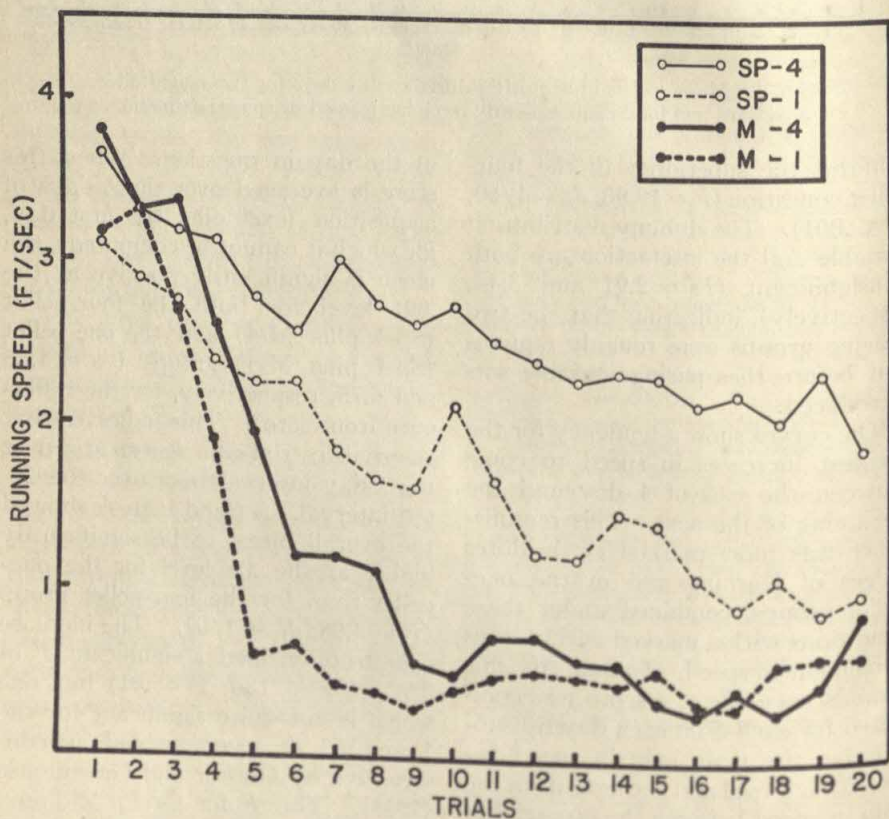


FIG. 2. Extinction speeds after two reward magnitudes with massed and spaced extinction.

magnitude using concentration of sucrose in the Skinner box (e.g., Collier & Willis, 1961; Guttman, 1953). They do not, however, explain the contradictory results of Armus (1959) and Hulse (1958).

The reminiscence effect in acquisition was unexpected. It is possible that this effect and the greater resistance to extinction in the distributed group may both be due to the same mechanism. This mechanism might be either reactive inhibition (Hull, 1951) built up during massed practice or, alternatively, activity drive (Hill, 1956) built up during rest in small cages and satiated by massed practice.

## Summary

Rats received 25 trials of acquisition and 20 trials of extinction in a straight alley, with reward magnitude (four pellets or one) and intertrial interval in extinction (20 sec. or 24 hr.) varied factorially. Resistance to extinction was greater for large reward and for spaced extinction, without the interaction predicted from a comparison of earlier studies. Marked reminiscence was observed from day to day in acquisition.

## References

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

Collier, G., & Willis, F. N. Deprivation and reinforcement. *J. exp. Psychol.*, 1961, **62**, 377–384.

Geisser, S., & Greenhouse, S. W. An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

Guttman, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

Hill, W. F. Activity as an autonomous drive. *J. comp. physiol. Psychol.*, 1956, **49**, 15–19.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

Hulse, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

Lewis, D. J. Acquisition, extinction, and spontaneous recovery as a function of percentage of reinforcement and intertrial intervals. *J. exp. Psychol.*, 1956, **51**, 45–53.

Metzger, R., Cotton, J. W., & Lewis, D. J. Effect of reinforcement magnitude and of order of presentation of different magnitudes on runway behavior. *J. comp. physiol. Psychol.*, 1957, **50**, 184–188.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

# HIERARCHIES IN CONCEPT ATTAINMENT

## ULRIC NEISSER AND PAUL WEENE [1]

### Brandeis University

In the laboratory or out of it, new ideas are always built on old ones. To attain a typical experimental concept, say "three borders," $S$ must already be able to identify borders, to count, to distinguish between $E$'s positive and negative statements, and so on. Much cognitive activity is hierarchically organized, in that the abstractions at one level form the basis of new abstractions at the next. The present experiment is an attempt to study hierarchical concepts explicitly. Although only binary concepts were used (more than two features were never relevant), the range of possibilities included three degrees of hierarchical depth.

There are many ways in which two or more features of a stimulus pattern may be combined into an attribute of higher order. For example, they may be conjoined: the attribute is defined by the joint presence of several features. A certain object is "of good quality" if it has been made skillfully (A), and of first-class materials (B). Neither feature alone is sufficient; both together are decisive. Bruner, Goodnow, and Austin (1956) worked extensively with conjunctive properties, but studied disjunctive attributes as well. In a disjunction, the presence of either property (or of both) is sufficient to define the concept. A patient may have an allergic reaction to either strawberries (A) or tomatoes (B). In conjunctive

concepts, the criterial attribute may be written symbolically as "A·B," while the corresponding notation for disjunctive attributes is "AvB." These two cases do not exhaust the possibilities, even if nothing matters but the presence or absence of two distinguishing features. There are 10 types of criterial attributes which can be based on one or on two features. They are listed in Table 1. Attributes based on more complex relations ("A followed by B," "A within B," and so on) will not be considered here.

The 10 types of bivariate attributes fall naturally into three levels, as indicated in Table 1. The univariate attributes are evidently the simplest. Next are a group of six bivariate attributes, made up directly from the univariate ones by negating, conjoining, or disjoining them. Finally, the two most complex attributes are formed by disjoining certain conjunctive pairs. Successive levels represent increasing complexity, not only in terms of the number of symbols needed to define the attributes, but in terms of a hierarchical structure. That is, attributes of Level II are combinations of those at Level I, and are components of those at Level III. It must be understood that this ordering arises only because we have taken negation, conjunction, and disjunction (rather than, say, double implication) as the basic operations, to be represented by elementary symbols. The hierarchy is merely a tautology until it is related to empirical findings like those presented here. In a sense, the findings of the present experiment support the selec-

TABLE 1

TYPES OF ATTRIBUTES WHICH CAN BE DEFINED BY PRESENCE
OR ABSENCE OF TWO FEATURES

| Name and Symbolic Designation | Description of Positive Instance | Example |
|---|---|---|
| **Level I** | | |
| Presence (A) | A must be present | Vertebrate: must have a backbone |
| Absence ($-A$) | A must not be present (complement of presence) | Invertebrate: must not have a backbone |
| **Level II** | | |
| Conjunction ($A \cdot B$) | Both A and B must be present | Good quality: both material and workmanship must be first class |
| Disjunction ($AvB$) | Either A or B or both must be present | Allergenic: a food which contains either tomatoes or strawberries (for example) |
| Exclusion ($A \cdot -B$) | A must be present and B not present | Eligible for Driver's license: must have passed test and not have committed felony |
| Disjunctive absence ($-Av-B$) | Either A or B, or both, must be absent (complement of conjunction) | Poor quality: either material or workmanship is not first class |
| Conjunctive absence ($-A \cdot -B$) | A and B must both be absent (complement of disjunction) | Nonallergenic: a food which contains neither tomatoes nor strawberries (for example) |
| Implication ($-AvB$) | A may be absent, but if A is present then B must be also; thus A implies B (complement of exclusion) | Ineligible for driver's license: must either have not passed test or have committed felony |
| **Level III** | | |
| Either/or ($A \cdot -B)v$ ($-A \cdot B$) | Either A or B must be present, but not both together | Negative product: either factor negative, but not both |
| Both/neither ($A \cdot B)v$ ($-A \cdot -B$) | Both A and B must be present, unless neither is (complement of either/or) | Positive product: both factors may be negative, or neither, but not just one |

tion of these three operations as primitive.

The 10 types of attributes fall into five complementary pairs. Anything that is a positive instance of one member of such a pair (i.e., which has its attribute) is a negative instance of the other. For example, $A \cdot B$ is the complement of $-Av-B$ because all and only those objects which are described by the former expression are not covered by the latter. Symbolically, one may find the complement of an expression by changing every "$\cdot$" to a "$v$" (and vice versa) and also every plus to a minus (and vice versa).

The experiment reported here is a study of the relative difficulty of attaining concepts at these several levels. The underlying hypothesis was that concepts at hierarchically higher levels would be more difficult to attain than those of lower levels.

## METHOD

*Experimental materials.*—The stimulus objects were strings of four consonants, each string printed on a 4 × 6 in. filing card. Only J, Q, V, X, and Z were used. Thus there were 625 distinguishable stimuli altogether (JJJJ, JJJQ, JJJV, ···, QJQZ, ···, VQZX, ···, ZZZZ). The concepts were defined in terms of the presence or absence of one or of two of these letters. The order and frequency of the letters in the string was never relevant, so that (for example) QQVZ was always

equivalent to VZZQ, ZVQV, and to any other string which contained Q, Z, and V but did not contain either J or X. Each of the types of criterial attribute represented in Table 1 could be realized in a number of specific ways. For example, JvX, QvZ, etc., are all disjunctions. It can easily be verified that altogether 110 different univariate and bivariate attributes can be defined on these stimuli. Any given stimulus is a positive instance of 55 of these and negative instance of the other 55. For example, QQVZ is a positive instance of Q, of $-X$, of VvJ, of $(Q \cdot V)v(-Q \cdot -V)$, etc., and a negative instance of $-Q$, of X, of $-V \cdot -J$, of $(Q \cdot -V)v(-Q \cdot V)$, etc.

*Subjects.*—The $S$s were 20 students of college age. They worked for about 3 hr. every morning, in groups of 5. A group of practiced $S$s could complete about four problems in such a session.

*Apparatus.*—The sequence of stimuli for a given concept was arranged as a deck of cards and set in a wooden frame, with the front card concealed by a spring-loaded shutter. To present a stimulus, $E$ released the shutter. Between trials, he closed the shutter and removed the front card.

*Procedure.*—When the stimulus appeared, each $S$, working independently, responded "plus" if he thought it was a positive instance of the attribute he was to discover, and "minus" if he thought it was not. Responses were made by means of toggle switches which controlled appropriate indicators on a panel visible only to $E$. When all $S$s had responded, $E$ noted the response, informed the $S$s of the correct answer, and then presented the next stimulus. No attempt to time the presentations was made, but an $S$ who hesitated more than about 15 sec. was asked to guess rather than delay further.

All sequences of stimuli were arranged to make positive and negative instances equally probable, and successive stimuli independent. (Appropriate sequences were prepared with an IBM 709 computer.) Since pure guessing would yield 50% correct responses, $S$ was judged to have attained a concept when he had made 25 consecutive responses with only a single error. (The possibility of carelessness made a 100% criterion inadvisable.) Ordinarily, a single problem was continued for 100 stimuli or until all $S$s had reached criterion. The situation was kept as noncompetitive as possible. The group was not informed about the performance of any individual, and each $S$ responded on every trial whether or not he had reached criterion.

Before the experiment, $S$s were told about the kinds of attributes that would be criterial. They were instructed that only the presence or absence of particular letters mattered, and that not more than two letters would be relevant. It was stressed that sequence and possible reduplication of letters on the cards was irrelevant. It was made clear that the absence of a letter, or of two letters, could be as important as its presence, and that the absence of one could be systematically connected with the presence of another.

*Experimental design.*—The first three problems (V, XvJ, Q · $-Z$) were the same for each group, and were considered practice. Explanation by $E$ of potentially relevant and irrelevant attributes continued during these problems. Thereafter, each group of 5 $S$s was given two consecutive cycles through the 10 types of problems described in Table 1. The order of problems within each cycle was varied from group to group, as were the letters which exemplified each type of concept; conjunction, for example, might be represented by J · X, Q · V, Z · J, etc. Thus each $S$ was presented with 23 concept attainment problems.

For two of the groups, a "nonresponding" cycle through the 10 types was interpolated between the three practice trials and the first of the cycles mentioned above. The $S$s were shown 100 positive instances of each concept, and then asked to write a description of it. Since these groups did not differ appreciably from the others in their performance on the concept-formation cycles, the data have been combined for this paper. The results of the nonresponding cycle (and of other such cycles carried out at the conclusion of the main experiment) were too ambiguous to merit description here.

## RESULTS

Table 2 exhibits the median trials needed to reach criterion on each type of problem, considering the two cycles separately. (Means cannot be given, because some $S$s failed to attain the criterion on some problems.) The results support the hypothesis that three distinct levels of difficulty are represented. Problems of Level II are systematically harder than those of Level I and easier than those of Level III. There is also a substantial practice effect: in 8 of 10 cases the

## TABLE 2
### Trials to Criterion for Different Types of Problems

| Type of Concept | Cycle 1 | | Cycle 2 | |
|---|---|---|---|---|
| | Median | $Q_1/Q_3$ | Median | $Q_1/Q_3$ |
| **Level I** | | | | |
| Presence (A) | 11.0 | 3.0/22.0 | 4.0 | 1.0/12.0 |
| Absence (−A) | 7.0 | 2.0/21.5 | 1.5 | 0.0/ 3.0 |
| **Level II** | | | | |
| Conjunction (A·B) | 13.0 | 6.0/43.5 | 18.0 | 4.5/50.5 |
| Disjunction (AvB) | 21.0 | 8.0/46.0 | 24.0 | 7.5/29.5 |
| Exclusion (A·−B) | 28.0 | 14.0/51.0 | 17.0 | 2.5/30.5 |
| Disjunctive absence (−Av−B) | 50.0 | 25.0/∞ | 23.0 | 9.5/37.5 |
| Conjunctive absence (−A·−B) | 29.0 | 17.0/61.0 | 8.0 | 3.0/18.0 |
| Implication (−AvB) | ∞ | 57.5/∞ | 19.5 | 9.0/59.0 |
| **Level III** | | | | |
| Either/or (A·−B)v(−A·B) | 68.0 | 47.5/∞ | 41.5 | 22.5/∞ |
| Both/neither (A·B)v(−A·−B) | ∞ | 54.5/∞ | 53.5 | 38.0/∞ |

Note.—$Q_1/Q_3$ indicates the first and third quartiles; $N = 20$ throughout. "∞" indicates that the median or quartile $S$ did not attain criterion. The 25 criterion trials are not included in these totals.

## TABLE 3
### Proportions of $Ss$ for Whom One Concept Was Easier Than Another: All Concept Pairs

| Level | Level III | | Level II | | | | | | Level I |
|---|---|---|---|---|---|---|---|---|---|
| | (A·B)v(−A·−B) | (A·−B)v(−A·B) | −AvB | −A·−B | −Av−B | A·−B | AvB | A·B | −A |
| **I** <br> A | 16/18* <br> 19/19* | 17/18* <br> 18/19* | 17/19* <br> 17/20* | 13/20 <br> 14/20 | 16/20* <br> 15/20* | 14/20 <br> 12/18 | 14/19 <br> 15/19* | 11/19 <br> 15/19* | 7/19 <br> 3/15* |
| −A | 18/20* <br> 19/20* | 18/19* <br> 19/20* | 18/19* <br> 19/20* | 16/18* <br> 15/18* | 17/20* <br> 17/18* | 14/20 <br> 14/17* | 14/17* <br> 16/19* | 12/18 <br> 17/19* | |
| **II** <br> A·B | 16/17* <br> 14/18* | 15/18* <br> 14/19 | 16/18* <br> 9/19 | 11/19 <br> 9/20 | 14/18* <br> 8/18 | 12/19 <br> 9/18 | 11/20 <br> 10/20 | | |
| AvB | 18/20* <br> 17/20* | 17/20* <br> 12/20 | 19/19* <br> 10/20 | 11/19 <br> 7/20 | 15/19* <br> 9/18 | 13/20 <br> 9/19 | | | |
| A·−B | 15/18* <br> 15/19* | 17/19* <br> 15/19* | 17/18* <br> 11/20 | 10/18 <br> 10/20 | 15/19* <br> 12/19 | | | | |
| −Av−B | 9/14 <br> 15/19* | 8/14 <br> 12/20 | 11/15 <br> 8/20 | 5/18 <br> 7/20 | | | | | |
| −A·−B | 14/17* <br> 19/20* | 15/19* <br> 19/20* | 15/16* <br> 14/19 | | | | | | |
| −AvB | 7/13 <br> 14/19 | 7/17 <br> 9/17 | | | | | | | |
| **III** <br> (A·−B)v <br> (−A·B) | 10/16 <br> 10/17 | | | | | | | | |

Note.—Each numerator is the number of $Ss$ who attained the concept of that row more quickly than the concept of that column; the denominator is the number available for the comparison. Upper fractions for Cycle 1, lower fractions for Cycle 2. Comparisons involving two different levels are above and to the left of the heavy line.
* $P \leq .05$; two-tailed binomial test.

median for the second cycle is below that for the first.

In Table 3, every type of concept is explicitly compared with every other type. The comparisons, made separately for the two cycles, are in terms of the proportion of $S$s who found one type easier than the other. Most proportions are based on slightly fewer than 20 $S$s, since those who found the two problems equally difficult, or solved neither, are not counted. For each comparison, the null hypothesis is that the two concepts are equally difficult, and that the tabulated proportion differs from $\frac{1}{2}$ only by chance. In all those cases where the comparison is between concepts of different levels, we have the counterhypothesis that an $S$ is more likely to find the lower-level hypothesis easier. Table 3 is so arranged that the counterhypothesis is supported by proportions above $\frac{1}{2}$, and not by those below $\frac{1}{2}$. It is also arranged so that all cases to which the counterhypothesis applies (i.e., comparisons between concepts at different levels) fall above and to the left of the heavy line. It appears that all but 1 of the 56 interlevel comparisons are in the predicted direction. Moreover, 39 of these proportions are significantly different from $\frac{1}{2}$ when considered individually. It is evident that levels of complexity play an important role in determining the difficulty of concept attainment.

No prediction was made about the relative difficulty of concepts within a single level. Indeed, Table 3 shows proportions near $\frac{1}{2}$ for most such comparisons. But implication $(-A \vee B)$ and disjunctive absence $(-A \vee -B)$ are significantly more difficult than the other second-level concepts on the first cycle of problems. The probable explanation is that $S$s did not fully understand the definition of these

concepts at first. On the second cycle this obstacle had been overcome by familiarity, and these concepts lost their special status. There is one other anomalous finding: $-A$ was easier than A. This result is difficult to understand, since these types differ only in which half of the universe of stimuli is called "plus."

## DISCUSSION

Why are higher-level concepts more difficult to attain? It might be supposed that, for complex combinatorial reasons, an unusually large number of stimuli is needed for logical elimination of competing hypotheses when a high-level attribute is the criterial one. We explored this possibility by writing a computer program (for the IBM 709) which solves our problems by rote. It has a list of the 110 possible concepts, and checks off those which are eliminated by each stimulus as it appears until only one concept remains. On the average, this program needs from 8 to 12 instances to pinpoint the defining attribute, although it may occasionally take much longer (if the string of stimuli happens to be unusually redundant). Paradoxically, the program takes slightly *longer* to identify the simple attributes (A and $-A$) than those of Level II, while the concepts of Level III take the fewest trials of all! The reason seems to be that when a series of stimuli are all compatible with a simple attribute such as "Z," there is a relatively high probability that they will all be compatible with certain high-level disjunctions, such as ZvQ, as well.

We wish to emphasize that the computer program was not written to simulate the behavior of human $S$s, but simply to establish the rates at which the different concepts could be attained by logical elimination. The discovery that human $S$s do not attain concepts in this way is hardly surprising.

A second explanation of the difficulty of attaining high-level concepts might appeal to the difficulty of formulating

them verbally. Perhaps *S*s find them unfamiliar, or cannot easily keep them in mind. The unexpected results with implication and disjunctive absence suggest that there is some validity to this interpretation. It is not fully adequate, however. The *S*s seemed to have a better verbal understanding of either/or than of most of the concepts at Level II which were more quickly attained.

In our opinion, higher-level concepts are more difficult because of their hierarchical organization. To identify an instance of $(Z \cdot Q) v(-Z \cdot -Q)$ one must have $Z \cdot Q$ and $-Z \cdot -Q$ available as components. After all, any individual instance of the first concept is also an instance of one of the latter two. Moreover, to work with $Z \cdot Q$, $S$ must know a $Z$ and a $Q$ when he sees one. Thus the levels into which we have divided the possible binary concepts may correspond to actual levels of input analysis by *S*s. To attain a complex concept, they must use, and therefore must have attained, preliminary concepts at lower levels.

## SUMMARY

Twenty *S*s were employed in a study of the relative difficulty of attaining 10 different types of concepts. All types involved only the presence or absence of two properties, but some were hierarchically more complex than others. For example, "Both A and B" is more complex than "A" but less complex than "Both A and B or neither." The results indicate that the difficulty of a concept varies directly with its complexity. This order of difficulty does not appear when a computer program is used to attain the concepts by simple elimination. It seems to reflect a hierarchical organization of conceptual processes in the *S*s themselves.

## REFERENCE

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

# REPLICATION REPORT: LATENT LEARNING IN A T MAZE AFTER SHOCK IN ONE END BOX

HENRY GLEITMAN AND MAGDALENA M. HERMAN

*Swarthmore College*

Tolman and Gleitman (1949) have reported latent learning in a T maze with highly differentiated end boxes. They found appropriate choice behavior after rats were shocked in one or the other of the two end boxes, following an equal number of reinforcements on both sides.

*Method.*—The original experiment was replicated in all respects but the following: (*a*) Guillotine doors were used instead of the one-way doors used in the original experiment, (*b*) all *S*s were run under 24 hr. of food deprivation, and (*c*) the location of the two differentiated end boxes was systematically varied, the dark one being on the right side for half the *S*s, and on the left for the other half. Finally, since there was some possibility that the positive results of the first experiment might be due to the distribution of trials (only two trials per day, one free and the other forced), two conditions of distribution were employed.

The *S*s were 38 experimentally naive female rats of white Angora strain, approximately 100 days old at the beginning of the experiment. All *S*s were reduced to 90% of their original body weight and kept at that level throughout the experiment. After 16 trials of pretraining on a straight runway, they were divided into two equal groups roughly matched on running times during pretraining. Group I received two trials per day on the apparatus over 10 days, the first trial free and the second forced. Group II received four trials per day over 5 days, the first and third being free and the others forced. One day following their last training trial, *S*s were placed in one of the two end boxes to find food, then into the other to receive two periods of intermittent shock. As in the original experiment, the spatial location of the end boxes was markedly different during this phase of the experiment as compared to training. Again as in the original experiment, half of the *S*s were shocked in the preferred, the other half in the nonpreferred end box. The *S*s were tested in the original apparatus, about 1 hr. after they had been shocked.

*Results.*—Fourteen out of 19 *S*s in Group I, and 13 out of 19 *S*s in Group II, chose the side away from that on which they had been shocked. It is thus apparent that at least for this limited range of values there was no effect of distribution of practice on the final choice. Since the two groups were virtually identical in their choice behavior, their results were combined and tested for statistical significance. Chance selection of the harmless side could be ruled out at the 1% level of significance ($CR = 2.60$, $P < .01$).

While the major finding of the original study was substantiated in the present experiment, there is some difference in the magnitude of the effects. In the present study, the harmless side was chosen by 71% of the *S*s, in the original experiment by 88%. This difference may be due to rather strong turning or place preferences developed in the course of the present experiment, which sometimes were strong enough to override other factors. Of the 11 *S*s who did *not* choose the harmless side of the final test, 9 were *S*s who had been shocked in the *preferred* end box.

## REFERENCE

TOLMAN, E. C., & GLEITMAN, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes followed by shock in one end-box. *J. exp. Psychol.*, 1949, **39**, 810–819.

(Received October 7, 1961)

# SUPPLEMENTARY REPORT: THE WEINSTOCK PARTIAL REINFORCEMENT EFFECT AND HABIT REVERSAL

LEON M. WISE

*Heidelberg College*

Sheffield (1949) found the partial reinforcement effect (PRE) for massed acquisition (15-sec. intertrial interval) but not for distributed acquisition (15-min. intertrial interval). Weinstock (1954, 1958) found the PRE under widely spaced trials (24-hr. intertrial interval). Both of the investigators used a simple running response. Wike (1953) and Grosslight and Radlow (1954) found the PRE for massed acquisition in a habit reversal discrimination problem. The purpose of the present experiment was to determine whether or not the PRE would be present in a habit reversal discrimination problem with a 24-hr. intertrial interval.

*Method.*—A 2 × 3 factorial design was used incorporating 100%, 70%, and 40% reinforcement, and 20-sec. and 24-hr. intertrial intervals. The $S$s were 60 experimentally naive male albino rats. A Y alley discrimination apparatus was employed. Stimuli and

trial procedures were essentially the same as in Grosslight and Radlow's experiment. All $S$s were given 40 acquisition trials and 40 habit reversal trials.

*Results and discussion.*—Figure 1 shows the mean number of correct responses for all groups for both acquisition and reversal. An analysis of covariance for the first five trials in massed habit reversal shows statistically significant differences among the 100%, 70%, and 40% groups ($F = 19.08$; $P < .01$) with the 100% group showing the least resistance to extinction (fastest reversal). Additional analyses of covariance at successive five trial intervals continue to show statistically significant differences in the same direction. This is in agreement with Sheffield's findings for massed acquisition. An analysis of covariance for the first five trials of distributed habit reversal shows no statistically significant differences among the three
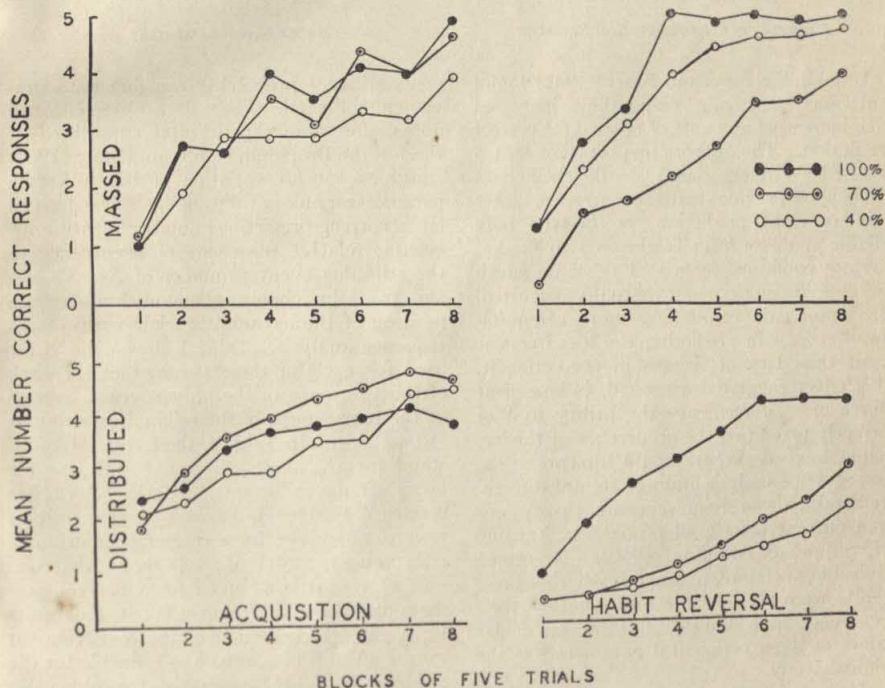


FIG. 1. Mean number of correct responses for all groups throughout acquisition and habit reversal in blocks of five trials.

groups.  However, a similar analysis conducted on the second five trials shows statistically significant differences ($F = 3.93$; $P < .05$) with the 100% group showing the least resistance to extinction.  Subsequent analyses conducted at successive five-trial intervals showed even greater statistical significance.  This finding does not agree with Sheffield's results for distributed acquisition. It does, however, substantiate the findings of Weinstock.

The present experiment can be added to a growing body of studies denying the Sheffield aftereffects hypothesis.  There seems to be little doubt now but that PREs can be obtained under both massed and distributed conditions and must be accounted for by any theory attempting to explain PREs.  Whether or not Weinstock's habituation hypothesis is

the correct interpretation the writer cannot say, but the present data are in agreement with it.

#### REFERENCES

GROSSLIGHT, J. H., & RADLOW, R.  Reinforcement schedules in habit reversal: A confirmation.  *J. exp. Psychol.*, 1954, **48**, 173–174.
SHEFFIELD, V. F.  Extinction as a function of partial reinforcement and distributed practice.  *J. exp. Psychol.*, 1949, **39**, 511–525.
WEINSTOCK, S.  Resistance to extinction of a running response following partial reinforcement under widely spaced trials.  *J. comp. physiol. Psychol.*, 1954, **47**, 310–322.
WEINSTOCK, S.  Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval.  *J. exp. Psychol.*, 1958, **56**, 151–158.
WIKE, E. L.  Extinction of a partially and continuously reinforced response with and without a rewarded alternative.  *J. exp. Psychol.*, 1953, **46**, 255–260.

(Received November 2, 1961)

## SUPPLEMENTARY REPORT: THE UTILITY OF CORRECTLY PREDICTING INFREQUENT EVENTS

YVONNE BRACKBILL              AND              ANTHONY BRAVOS

*University of Colorado Medical School*              *Johns Hopkins University*

Brackbill, Kappy, and Starr (1962) found that maximum gain responding increased with increasing amounts of reward for correct prediction.  The authors' expectation that a first-order sequence analysis of their data would show previous actual occurrence, rather than previous prediction, to be the only reliable predictor from Trial $n - 1$ to Trial $n$, was not confirmed for $n - 1$ trials on which the less frequent event actually occurred. Maximum gain responding more often followed success in predicting the less frequent event than lack of success in predicting it. This effect suggested a second, independent source of reinforcement—the utility to $S$ of correctly predicting the occurrence of the less frequent event.  Whatever the interpretation, the sequence analysis findings are not directly predictable from reinforcement theory nor from current theories of probability learning (cf. Suppes & Atkinson, 1960).  It seemed advisable, therefore, to find out whether these results were reproducible and whether their occurrence was limited to the particular values of the experimental parameters of the original study.

*Method.*—First-order sequence analyses were performed on 12 independent sets of

noncontingent probability learning data previously collected.  These data were obtained under the same experimental conditions as those of the Brackbill, Kappy, and Starr (1962) study except for variation of the following parameters: amount of tangible reward given for a correct prediction; number of stimulus events; relative frequency of occurrence of the stimulus events; number of $Ss$; $S$'s age and grade in school; and number and series position of the asymptotic trials within each sequence analysis.  Table 1 shows the value used for each of these parameters for each of the 12 groups of the present study as well as the four groups of the original experiment (Rows 2–5).  In Table 1, the letters M and L stand for the more (or most) and less (or least) frequent events.  Under "tangible reward," 1 M or L shows that one unit of reward was given for a correct prediction of either event, and 1 M: 4 L shows that one unit of reward was given for a correct prediction of the more frequent event, four units for a correct prediction of the less frequent event.  A unit of reward was 1 marble for the younger $Ss$ and 1 point for the older $Ss$; 100 marbles were exchanged for one toy, and 100 points for $1.00.  In the last five rows of

TABLE 1

SUMMARY OF SUCCESSIVE TRIAL CONTINGENCIES FOR 16 SETS OF PROBABILITY LEARNING DATA
OBTAINED UNDER VARYING EXPERIMENTAL CONDITIONS

| Tangible Reward | No. and Relative Frequency of Occurrence of Stimulus Events | $S$s | | No. and Position of Trials Analyzed | Mean Probability of Predicting Event M on Trial $n$, Given Prediction (p) and Occurrence (o) on Trial $n - 1$ | | | |
|---|---|---|---|---|---|---|---|---|
| | | $N$ | School Grade | | $M_pM_o$ | $L_pM_o$ | $M_pL_o$ | $L_pL_o$ |
| None | 75 : 25 | 4 | 5 | 321–400 | .77 | .71 | .56 | .68 |
| None | 75 : 25 | 12 | 3 | 101–200 | .74 | .82 | .38 | .56 |
| 1 M or L | 75 : 25 | 12 | 3 | 101–200 | .80 | .83 | .65 | .85 |
| 3 M or L | 75 : 25 | 12 | 3 | 101–200 | .82 | .87 | .67 | .73 |
| 5 M or L | 75 : 25 | 12 | 3 | 101–200 | .89 | .87 | .64 | .78 |
| 1 M:4 L | 75 : 25 | 10 | 4 | 121–200 | .76 | .79 | .50 | .75 |
| 1 M:3 L | 75 : 25 | 10 | 4 | 121–200 | .83 | .70 | .68 | .76 |
| 2 M:3 L | 75 : 25 | 10 | 4 | 121–200 | .80 | .84 | .68 | .85 |
| 1 M:4 L | 75 : 25 | 10 | 12 | 121–200 | .74 | .57 | .71 | .61 |
| 1 M:3 L | 75 : 25 | 10 | 12 | 121–200 | .72 | .62 | .63 | .76 |
| 2 M:3 L | 75 : 25 | 10 | 12 | 121–200 | .84 | .88 | .70 | .72 |
| None | 40: (35): 25 | 10 | 6 | 301–400 | .47 | .48 | .27 | .31 |
| None | 50: (30): 20 | 9 | 6 | 201–400 | .60 | .63 | .22 | .48 |
| None | 60: (25): 15 | 8 | 6 | 201–400 | .76 | .68 | .29 | .42 |
| None | 70: (20): 10 | 10 | 6 | 301–400 | .71 | .77 | .36 | .50 |
| None | 80: (15): 5 | 10 | 6 | 301–400 | .86 | .72 | .67 | .60 |

Column 2, the frequencies in parentheses indicate that Table 1 does not include the sequence analysis results for the stimulus events of intermediate frequency under the three-stimulus conditions.

*Results and discussion.*—The last four columns of Table 1 show the mean probabilities of predicting Event M on Trial $n$, given the prediction (p) and actual occurrence (o) on Trial $n - 1$. Thus, for example, the entry in the upper right-hand cell indicates that, for those instances in which $S$s had predicted the less frequent event ($L_p$) on Trial $n - 1$, and the less frequent event had actually occurred ($L_o$) on Trial $n - 1$, the mean probability of predicting the more frequent event (M) on Trial $n$ was .68.

The question under investigation is whether $S$'s prediction on Trial $n$ is determined by the nature of his previous prediction as well as by the previous actual occurrence or reinforcement. Therefore, it is appropriate to compare the $M_pM_o$ to the $L_pM_o$ probabilities and the $L_pL_o$ to the $M_pL_o$ probabilities. For the present data, shown in Rows 1 and 6–16 of Table 1, the mean value of $M_pM_o$ exceeds that of $L_pM_o$ in 6 cases out of 12, while the mean value of $L_pL_o$ exceeds that of $M_pL_o$ in 10 cases out of 12 ($P = .04$, by

binomial expansion). For all 16 sets of data, the mean value of $L_pL_o$ exceeds that of $M_pL_o$ in 14 cases ($P = .004$).

In spite of wide variations within several experimental parameters, the same result has emerged as before. In order to maximize prediction to Trial $n$ from preceding trials on which the less or least frequent event occurred, it is necessary to consider $S$'s previous prediction in addition to the previous actual occurrence. Also, the direction of the effect in the present results supports the original interpretation: that there is a relatively greater utility to $S$ of correctly predicting the occurrence of the less (or least) frequent event. It would be interesting to see if the same phenomenon might occur generally in *any* type of learning situation in which $S$, finding $E$'s "game" tedious and uninteresting, can and does invent one of his own.

#### REFERENCES

BRACKBILL, Y., KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

(Received November 16, 1961)

# SUPPLEMENTARY REPORT: FREQUENCY OF STIMULUS PRESENTATION AND SHORT-TERM DECREMENT IN RECALL [1]

## S. HELLYER

*Defence Research Medical Laboratories, Toronto, Canada*

Peterson and Peterson (1959) report clear-cut evidence of a progressive improvement in recall scores with an increase in the number of repetitions of the material by $S$ before the delay of recall began. Certain anomalies in their data suggested the obtained differences might have resulted from $E$ unintentionally interfering with $S$'s response pattern. The present experiment repeated the Petersons' work using visually presented material to reduce the likelihood of inadvertent interference.

*Method.*—The verbal items were three-consonant units with a Witmer association value no greater than 33%. The material used to keep $S$s active during the recall delay interval consisted of groups of three randomly selected digits.

One three-consonant unit and a series of digit groups were typed as a list on a memory drum tape. There were eight such lists on a tape and a pair of tapes constituted a set of all 16 experimental conditions in random order, i.e., one, two, four, and eight repetitions of the three-consonant unit and recall delay intervals of 3, 9, 18, and 27 sec. The stimuli were displayed at a rate of 1/sec.

Five seconds after $E$ started the memory drum, a green star appeared in the window as a warning that the consonant unit was about to appear. The $S$ was instructed to read aloud what appeared in the window and not to anticipate what might appear next. This was done to decrease the possibility that $S$ was preparing for another rehearsal as the three digits of the intervening activity appeared. When the entire list had been presented, a red star appeared as a signal to recall the consonants presented at the start of the list. The intervening activity consisted in reading groups of three digits. After recall of the consonants was completed, $S$ was required to make two judgments about these numbers, estimates as to which digit had appeared least frequently and which most frequently. This was done to make the number task a more meaningful part of the experiment. On each of the 5 days, a

### TABLE 1
#### PROPORTIONS OF ITEMS CORRECTLY RECALLED

| Number of Presentations | Recall Delay Interval (Sec.) | | | |
|---|---|---|---|---|
| | 3 | 9 | 18 | 27 |
| 8 | .99 | .89 | .74 | .66 |
| 4 | .94 | .73 | .56 | .46 |
| 2 | .92 | .54 | .31 | .22 |
| 1 | .89 | .38 | .21 | .14 |

different pair of lists was presented in a random order to each $S$. The 25 paid $S$s were housewives.

*Results and discussion.*—Following Peterson and Peterson (1959), an item was considered to be correctly recalled only if every consonant was correct and in its proper position. Table 1 records the mean proportion of items recalled correctly for 25 $S$s on 5 days.

These data confirm the Petersons' conclusions that there is better recall with an increase in the number of stimulus repetitions and with shorter periods of delay before recall.

An analysis of variance showed that number of presentations and recall delay interval are both significant ($P < .01$). The only significant interaction was recall delay with number of presentations. This arises because the effect of an increase in recall delay was more pronounced on the trials where the consonant groups were presented once or twice than when they were presented four or eight times.

The $S$s in the present experiment obtained markedly higher recall scores than those reported by Peterson and Peterson, perhaps because in the present study the stimuli were presented both visually and aurally. Inspection of the present data, confirmed by an analysis of variance, yields no evidence for learning over blocks of trials.

### REFERENCE

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.,* 1959, 58, 193–198.

(Received November 25, 1961)